



Spatially Structured Sparse Morphological Component Separation for Voltage-Sensitive Dye Optical Imaging

Hugo Raguet, Cyril Monier, Luc Foubert, Isabelle Ferezou, Yves Fregnac, Gabriel Peyré

► To cite this version:

Hugo Raguet, Cyril Monier, Luc Foubert, Isabelle Ferezou, Yves Fregnac, et al.. Spatially Structured Sparse Morphological Component Separation for Voltage-Sensitive Dye Optical Imaging. *Journal of Neuroscience Methods*, 2016, 257, pp.76-96. 10.1016/j.jneumeth.2015.09.024 . hal-01200646

HAL Id: hal-01200646

<https://hal.science/hal-01200646>

Submitted on 16 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatially Structured Sparse Morphological Component Separation for Voltage-Sensitive Dye Optical Imaging

Hugo Raguet^a, Cyril Monier^b, Luc Foubert^b, Isabelle Ferezou^b, Yves Fregnac^b, Gabriel Peyré^a

^aCNRS and Ceremade, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775, Paris Cedex 16, France

^bUnit of Neuroscience, Information and Complexity, CNRS UPR-3293, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

Abstract

Background. Voltage-sensitive dye optical imaging is a promising technique for studying *in vivo* neural assemblies dynamics where functional clustering can be visualized in the imaging plane. Its practical potential is however limited by many artifacts.

New Method. We present a novel method, that we call “SMCS” (Spatially Structured Sparse Morphological Component Separation), to separate the relevant biological signal from noise and artifacts. It extends Generalized Linear Models (GLM) by using a set of convex non-smooth regularization priors adapted to the morphology of the sources and artifacts to capture.

Results. We make use of first order proximal splitting algorithms to solve the corresponding large scale optimization problem. We also propose an automatic parameters selection procedure based on statistical risk estimation methods.

Comparison with Existing Methods. We compare this method with blank subtraction and GLM methods on both synthetic and real data. It shows encouraging perspectives for the observation of complex cortical dynamics.

Conclusions. This work shows how recent advances in source separation can be integrated into a biophysical model of VSDOI. Going beyond GLM methods is important to capture transient cortical events such as propagating waves.

Keywords: voltage-sensitive dye optical imaging, sensory cortical dynamics, orientation maps, sparse component separation

1. Introduction

This work is focused on the processing of *voltage-sensitive dye optical imaging* (VSDOI) recordings. This experimental method can in theory, combine the mesoscopic spatial scale of optical imaging with the real-time temporal resolution of direct electrophysiological measurements. The VSDOI signal is however contaminated by a strong noise and artifacts, and suffers from a poor understanding of the signal formation process, as we detail in this section.

1.1. The VSDOI Experimental Method

1.1.1. Principles

Sensitivity to membrane potential of some fluorescent dyes is known for long (at least back to [Cohen et al. \(1974\)](#)), and its use for monitoring neuronal activity has been considered ever since. Starting with recordings of action potentials in individual neurons ([Davila et al., 1973](#); [Salzberg et al., 1973](#)), the technique got continuously improved along time, and is now a privileged experimental method for monitoring cortical activity simultaneously at several locations with both high spatial and temporal resolution ([Grinvald and Hildesheim, 2004](#)).

To perform VSDOI one stains the neuronal tissue with voltage-sensitive dye, which are fluorescent molecules, also called *fluorophores*. Some of those bind to neuron membranes. Each fluorophore, when illuminated with the correct exciting wavelength, emits light in a different wavelength in return. The fluorophores that are bound to a neuron membrane happen to emit differently according to the electrical potential at the membrane. Thus, recording the emitted fluorescence provides access to the variations of the potential of those membranes along time. See for instance [Grinvald et al. \(1999\)](#) or [Frostig \(2009\)](#) for detailed practical methodologies of *in vivo* recordings.

1.1.2. Limitations

Because the changes in fluorescence closely follow a change in membrane potential, both in time (order of the microsecond) and in space (at molecular scale), the resolution of VSDOI is in theory only limited by two main factors: the quantum nature of photon emissions, and the precision of the optical recording device.

Recorded fluorescence intensity corresponds to the count of the number of photons reaching the detector during a certain time laps. At a given intensity, the number of emitted photons during a given duration is best modeled by a Poisson distribution ([Foschini et al., 1975](#)), for which the *variance* increases linearly to the mean level. Hence, the expected quality of a fluorescence measurement, quantified by the ratio between the mean number of emitted photons and its standard deviation, is proportional to the *square root* of the mean number of photons. As a consequence, it is important to maximize the mean num-

*Corresponding author: Gabriel Peyré

Email addresses: hugo.raguet@gmail.com (Hugo Raguet),
cyril.monier@unic.cnrs-gif.fr (Cyril Monier),
luc.foubert@unic.cnrs-gif.fr (Luc Foubert),
isabelle.ferezou@unic.cnrs-gif.fr (Isabelle Ferezou),
yves.fregnac@unic.cnrs-gif.fr (Yves Fregnac),
gabriel.peyre@ceremade.dauphine.fr (Gabriel Peyré)

ber of photons recorded at each measurement. This is in turn proportional to three main factors: the number of fluorophores within the focus of the detector, the duration of each measurement, and the intensity of the exciting light. In any experimental set-up, a compromise must be found between those three factors, in accordance with the instrumental and material conditions.

On the one hand, the quantity of fluorophores reporting potential changes is limited by the total area of neuronal membrane under investigation, and by the physical access to those membranes. Moreover, the quantity of fluorophore introduced in the medium must also be limited in order to avoid *pharmacological* effects, *i.e.* perturbations of the functioning of the neurons due to the presence of the fluorophores. Similarly, the intensity and the duration of the exposure of the exciting light is limited by *photodynamic* damage that can occur to the medium. See for instance Grinvald et al. (1999) and Peterka et al. (2011) for more details on those practical limitations.

On the other hand, optical precision is limited by *light scattering* and focal precision, while the quantity of data recorded during a certain amount of time is limited by acquisition and data-storage speed. Horizontal spatial resolution (parallel to the cortical surface and orthogonal to the optical axis) goes below a tenth of a micron when using confocal (Holthoff et al., 2010) or two-photon (Acker et al., 2011) microscopy. This actually allows discriminating between neuron compartments. Unfortunately, only two-photon techniques allow such precision in the vertical axis. For all other recording techniques, the focal-depth is much thicker. This can be limited *in vitro* by reducing the thickness of the sample itself (thin brain slices), but *in vivo*, contributions of several cortical layers might be mixed in the signal, depending on the dye distribution along the depth of the cortex, see for instance (Ferezou et al., 2006; Lippert et al., 2007; Chemla and Chavane, 2010a). Concerning the spatial extent of the field of view, most commercial devices available nowadays exceed 1000×1000 pixels, and temporal resolution goes up to 10 kHz, allowing to capture every events of an action potential (Tominaga and Tominaga, 2013).

The final compromise between spatial and temporal resolution, data quantity, and recording quality depends on the phenomena under investigation and on the experimental conditions. In particular, *in vitro* experiments allow more flexibility than *in vivo*, but do not give access to the same information. In all the present work we are mostly interested in *in vivo* recordings, at population level, with spatial resolution between 10 and 100 μm , sampling frequency between 100 Hz and 1 kHz, and spatial extent no more than 100×100 pixels.

1.2. The Targeted Signal

Before diving into the technical aspects of VSDOI signal processing, let us describe the phenomena one would like to investigate thanks to the VSDOI experimental method. Those considerations are useful in order to understand the motivation and ambition that lies behind the work developed in this manuscript.

1.2.1. The Cortical Phenomena Under Investigation

The use of VSDOI at single cells level is justified when direct intracellular recording is rendered impossible, often because it is too invasive or because the targeted site is too small for the insertion of an electrode. The advantage of the VSDOI experimental method which is the most interesting to us is its ability to record in real-time the activity of entire networks comprising thousands of neurons. At the population level, VSDOI is more sensitive to subthreshold potential variations of many synchronous neurons than to individual spiking activity (Chemla and Chavane, 2010b). Although it does not reveal the action potentials, such mesoscopic information is useful to understand the mechanisms of integration of individual neurons activity within local networks. In particular, spatiotemporal dynamics of functional structures, population encoding of sensory stimuli, long-range connectivity, and propagating phenomenon can be studied; see Grinvald and Hildesheim (2004) and Chemla and Chavane (2010b) for more details on the cortical mechanisms best revealed by VSDOI.

1.2.2. The Question of the Ongoing Activity

When studying functional organization of the cortex, the usual approach is to analyze the activity evoked *in vivo* by certain stimuli. However, it has been observed that the variability of the neuronal response to several repetitions of the same stimulus is sometimes as large as the response itself; see Arieli et al. (1995a) and references therein. Moreover, even in the absence of specific stimulus, neurons exhibit spontaneous activity that is often highly structured in space and time, at both at the single cell level and at the population level. This spontaneous or *ongoing* activity can be observed with any recording method, and presents a wide variety of dynamics, usually highly dependent to the overall state of the subject (anesthetized, awake or attentive).

Many important questions arise about the ongoing activity. Neither its origin and mechanisms, nor its relationship to the activity evoked by a specific stimulus and its role in perceptual attention are yet well understood. Experiments with VSDOI could provide precious information, but let us emphasize here that ongoing activity is also an obstacle for the single trial analysis of VSDOI recordings at population level. As described in the next section, many nonneuronal artifacts corrupts VSDOI acquisitions, so that up to now information is extracted by averaging over repetitions, or using ad-hoc processing methods retrieving the evoked, reproducible signal and discarding the variability. A notable exception is Muller et al. (2014), who shows that single trial analysis is important to detect propagating waves of activity. Although VSDOI has already been used in studies of ongoing dynamics (see again Arieli et al. (1995a), or Arieli et al. (1996)), precaution must be taken for their interpretation. This underlines the need for new processing methods that would capture all the variability at the single acquisition (also dubbed *trial*) level.

On a VSDOI signal processing point of view, it is important to distinguish several temporal and spatial scales of ongoing dynamics. Various dynamical patterns have been reported in

the literature, depending on the local or global origin of synchronous events in the network. Some of these events take the form of propagating waves across the laminar dimension of the cortical network. Others act more as standing waves of “up” and “down” states where all neurons simultaneously engage in a high conductance state or on the contrary return to their resting state. These “up” and “down” states are respectively more *depolarized* (high conductance) and more *hyperpolarized* (resting state with reduced synaptic bombardment) temporary states, resulting respectively in higher and lower activity levels for durations in the order of the second. Such slow fluctuations are usually associated with high level of synchrony, involving entire networks at the spatial scale of the millimeter, see for instance [Lampl et al. \(1999\)](#); [Petersen et al. \(2003\)](#).

Then, faster spontaneous events that are often reported concern propagating waves of activity running across neuronal networks, especially in the cortex. Such events are highly structured, and usually closely resemble events that can be evoked by specific stimuli, see [Contreras \(2007\)](#); [Xu et al. \(2007\)](#); [Benucci et al. \(2007\)](#). They are however very diverse, with many different propagation velocities and spatial extents, with frequency scales in order of magnitude from 1 up to 100 Hz, see for instance the review of [Muller and Destexhe \(2012\)](#). Finally, individual neurons within networks always exhibit fluctuations of activity, both in term of spiking activity and of subthreshold potential variations, due to a wide variety of sources. There are often described as random noise (review in [Muller and Destexhe \(2012\)](#)) although spectral analysis reveals colored correlation in the synaptic background activity [Boustani et al. \(2009\)](#). Notably, these fluctuations might have characteristic frequency above 100 Hz and show little correlation from one neuron to another.

This division according to three different spatiotemporal frequency scales might, or might not, have actual biological relevance. However, they seem to relate to distinct phenomena and, more importantly, their respective influences on VSDOI acquisitions are as significant as they are different.

1.3. The Challenge of VSDOI in Vivo

In spite of the constant enhancement of the VSDOI technique along several decades, *in vivo* recording remains a technical challenge, due to a wide variety of artifacts and noise that corrupt the signal, and that can also differ greatly according to the experimental conditions.

1.3.1. From Fluorescence to Neuronal Activity: What Are We Recording?

As introduced in § 1.1.1, the fluorophores bound to a neuronal membrane fluoresce differently according to the electrical potential at the membrane. This knowledge, however, is not sufficient to deduce changes in neurons membrane potential. First of all, hundreds to thousands of neurons contribute to each single pixel of the recorded acquisition and it is impossible to differentiate between contributions of axonal or dendritic activity, of inhibitory or excitatory neurons, or even of some non-neuronal cells (glia, etc.). A realistic biophysical model of the

different contributions has been proposed by [Chemla and Chavane \(2010a\)](#). In general, the best information available is the *variation of membrane potential averaged over multiple compartments of multiple cells, integrated over several cortical layers*. Moreover the interpretation of this variation of membrane potential is delicate because one does not know the *baseline activity*, the reference value to which the variation should be computed.

Now, how does one link the VSDOI acquisition to this average potential variation? The mechanism linking variations in membrane potential to variations in fluorescence has already been investigated ([Peterka et al., 2011](#)) and the relationship between those variations is supposed to be linear. But what is the *gain* of this linear relationship? For a given spatial position and at a given instant (a pixel of a frame in the acquisition), it should depend on the *quantity of fluorophores bound to the membrane* and on the illumination intensity. Here again one often assumes linear relationship between intensity of re-emitted fluorescence and fluorophore concentration, see for instance [Tanke et al. \(1982\)](#) in the context of microscopy. Also, keep in mind that only fluorophores bound to a neuronal membrane contribute to the desired signal, so that the area of stained membrane should be taken into account as well. Unfortunately one does not have access to this information.

Moreover, VSDOI records the *absolute* fluorescence intensity (*i.e.* the number of photons detected during the frame duration at each pixel). In order to get variations of fluorescence, one needs another baseline value, the *baseline fluorescence*, that is to say the fluorescence intensity that would be recorded independently from any neuronal activity (and from other existing biophysical sources of variations). Baseline activity and baseline fluorescence are hard to distinguish from the sole information of the recorded fluorescence. This is an important issue in VSDOI, because many studies could be influenced by the way baseline activity is defined and estimated. Indeed, recall from § 1.2.2 that ongoing activity can reach the order of magnitude of an evoked activity, so that it is impossible to define the baseline activity as the level of activity when no stimulus is presented. It is common to estimate the baseline activity mixed with the baseline fluorescence, as the recorded values in the first frames of the acquisition where no evoked activity is expected. It is then subtracted from all the frames (see § 2). Using blank subtraction, influence of spontaneous activity over those first frames is passed on to the entire acquisition, and should be taken into account. Unfortunately, this is difficult: variations of fluorescence in VSDOI that is actually due to neuronal activity is known to be in the order of a *thousandth* of the baseline fluorescence ([Grinvald et al., 1999](#)). Even worse: many non-neuronal sources of variations actually *exceed* this ratio.

1.3.2. Sources of Noise and Artifacts

Many phenomena cause measurement errors. Like every fluorescence based experimental method, VSDOI is affected by *dye photobleaching*. When illuminated by exciting light, the fluorophores have the tendency to degrade or to react with other molecules so that they do not emit light anymore ([Song et al., 1995](#)). This results in an overall decrease of the recorded flu-

orescence over time, independently from any neuronal activity. It is important to note that several mechanisms are involved and bring forth different bleaching dynamics with distinct time constants. For instance, when fluorophores form nonfluorescent complexes with other molecules in a reversible fashion, they will fluoresce again when they retrieve their initial conformation. This results in a *fast* fluorescence decrease affecting the signal (as observed for instance in [Chen et al. \(2008\)](#); [Reynaud et al. \(2011\)](#)) as soon as the observation field is illuminated for recording, until a steady-state equilibrium is reached. Switching off the illumination modifies the equilibrium back to the previous state and the concentration of fluorescing molecules progressively goes back close to its initial value. Now, if a fluorophore degrades, it stops to fluoresce permanently. This process causes a *slow* fluorescence decrease (as reported for instance in [Lippert et al. \(2007\)](#); [Takagaki et al. \(2008\)](#)), hardly detectable at the time scale of a single acquisition. However, it can be seen by comparing successive acquisitions recorded during several hours of experimental protocol, until the fluorescence is too low for recording. The term bleaching or photobleaching is used independently in the literature to refer to both phenomena; and even sometimes to other phenomena that are strictly speaking unrelated to photobleaching but lead to similar fluorescence dynamic, *e.g.* wash-out of the dye along experimentation (see again [Lippert et al. \(2007\)](#); [Takagaki et al. \(2008\)](#)).

For *in vivo* experiments, emitted fluorescence due to neuronal activity suffers from other important perturbations. This is mainly due to absorption properties of the biological tissues and of the hemoglobin in the blood of the living animal, as described extensively in [Shoham et al. \(1999\)](#). In particular most *in vivo* VSDOI acquisitions are contaminated by periodic components corresponding to the heartbeats of the animal; similarly, respiration artifacts are present. Also, the intrinsic signal, which is actually the signal of interest when performing intrinsic imaging, causes in VSDOI an intrinsic artifact. On acquisitions longer than one second, it leads to a slow decrease followed by an increase of the fluorescence subsequently to high neuronal activity.

Finally, we call *shot noise* the quantum fluctuations of the exciting light and of the emitted fluorescence described in § 1.1.2. Let us also mention mechanical vibrations of the experimental table, oscillations of the alternative current that supplies the experimental device, and thermal noise of the camera as other possible nonneuronal sources of fluorescence variations ([Grinvald et al., 1999](#)). Those phenomena create high frequency fluctuations of the VSDOI signal, which are difficult to separate from high frequency ongoing activity.

1.3.3. From Neuronal Activity to Fluorescence: Modeling the Signal

Even though artifacts present distinctive dynamics, the way they act on the signal is nontrivial. Supposing for instance that the bleaching affects uniformly all fluorophores in the medium, its action on the signal should be strictly multiplicative, that is to say bleaching is a modulation of the gain over time. Alternatively if bleaching affects mostly fluorophores that are *not*

bound to a neuronal membrane, its influence with respect to the neuronal signal should be considered additive. Interestingly, those two opposed alternatives are considered in the literature without further motivation (see [Bonhoeffer and Grinvald \(1996\)](#) for a discussion in the case of intrinsic imaging), for instance when performing blank division and blank subtraction methods (see § 2.1). Note however that since the baseline fluorescence dominates all other components in the signal, both assumptions often lead to similar results, so that to our knowledge, the question has never truly been discussed. Note also that in most processing techniques, all contributions in the signal are considered additive (see § 2).

A similar question arises about the nature and statistics of random noise. On the one hand, noise that results from a multitude of small contributions, like random fluctuations of membrane potential, are usually modeled as Gaussian noise ([Reynaud et al., 2011](#)), with neglected spatiotemporal correlations. On the other hand, recall from § 1.1.2 that shot noise is best modeled with a Poisson distribution, for which the variance increases proportionally to the mean signal level. Note once again that since the constant baseline fluorescence dominates all other components in the signal, this can be neglected when analyzing the time course of a single pixel. However important variations of the baseline fluorescence from one pixel to another can lead to significant bias when performing spatial analysis without taking it into account.

2. Previous Works

Depending on the acquisition material (type of dye, camera and other device-specific conditions), on the observed cortical network (which animal, anesthetized or awake, which cortical area), on the observed phenomenon (spontaneous or evoked activity), on the length of the acquisitions (from less than one second to several seconds), and on the duration of the protocol (up to several hours of recording), the possible artifacts and noise do not have the same influence on the observed signal. This is why it exists in the literature many different approaches to extract the neuronal activity of interest from the raw signal.

We make here a (nonexhaustive) list of those approaches, in order to illustrate their variety and to infer some common *priors* that are made on the nature of the components in the VSDOI signal.

2.1. Blank Subtraction

The most common processing is the *blank subtraction* (BkS). A blank is an acquisition recorded while no stimulus is presented to the animal. With an acquisition in stimulus condition (called stimulus acquisition in the following) and a corresponding blank, one retrieves the stimulus evoked neuronal response by essentially subtracting the latter from the former. Usually both acquisitions are first separately normalized by dividing each frame (pixel-to-pixel) by a so-called *zero-frame* which is the mean of several frames taken at the beginning of the acquisition. The blank subtraction method relies on the assumption that the artifacts and noise act additively on the signal, and that

their time courses are the same in the blank and in the stimulus acquisitions. It is thus better to record a corresponding blank just before or right after any stimulus acquisition, so that the conditions are as similar as possible. This is significantly improved by synchronizing the acquisitions with the heartbeat of the animal, and even with its respiration in the case of anesthetized animals under intubation. The division by the zero-frame assumes that all the components are proportional to the same resting fluorescence value which can be estimated on the first frames. The division pixel-to-pixel accounts for the differences of fluorophore concentration and of illumination from one spatial position to another. Let us finally note that simple subtraction assumes that the baseline fluorescence of both the blank and the stimulus are the same (after zero-frame division).

Variations on this method can be found. The blank is sometimes smoothed by low-pass filtering or by averaging all blank acquisitions (Markounikau et al., 2010). It can also be replaced by the so-called “cocktail blank”, which is the mean over a set of acquisitions supposed to active successively all regions of the observed cortical area, in an uniform fashion (Grinvald et al., 1999). The advantage of using a “cocktail blank” reference is to remove the important stimulus-non-specific response (characterizing typically VSDOI signal) that shows an amplitude as large as one order of magnitude compared to the stimulus-specific-modulations. The disadvantage of removing the non-specific-response is the loss of information about the overall neural dynamics that might be carried into this common modulation signal. Also if a notion of complementarity can be defined over the space of stimuli (e.g. when studying ocular dominance in the cortical domain or orientation selectivity in the visual cortex as a function of particular stimuli features) it is possible to subtract an orthogonal stimulus acquisition instead of subtracting a blank (Shoham et al., 1999). This does not give access to the evoked neuronal dynamic but it is useful for revealing differential sensitivity or simply enhancing contrast in functional preference in cortical maps. Sometimes a division is considered in place of a subtraction (Grinvald et al., 1999; Jancke et al., 2004). As discussed in § 1.3.3, there is no consensus about which approach is best.

2.2. Baseline Fluorescence, Gain and Bleaching

Chakraborty et al. (2007) perform VSDOI *in vitro* (on brain slices) so that the signal is not affected by physiological artifacts. They evaluate both baseline fluorescence and gain by averaging the first 100 ms of each acquisitions for each pixel. The (fast) bleaching is then handled by subtracting to the time course of all pixels a single linear function fitted on the mean over the entire image. To avoid frames where the evoked response is the most likely to appear, the fit is performed only on the second temporal half of the acquisition. A similar method is used in Chen et al. (2008) but on spatial bins of pixels. The authors do not describe how they select the bins. They evaluate both a baseline fluorescence and a gain for each bin by averaging the signals over all time frames of all trials, even though slow bleaching can affect baseline fluorescence and gain from one trial to the other. Then a linear function is fitted and sub-

tracted for each bin. The fit is performed on the time frames before and after the expected response period.

As already mentioned, some authors reporting on VSDOI observe slow bleaching (*i.e.* with a time constant between minutes and hours, see § 1.3.2) but no fast bleaching (*i.e.* with time constant less than a second). Our guess is that they focus on long recordings (10 s or more) for which the first seconds of illumination are not recorded or used for the analysis. Concerning the slow bleaching affecting the gain and the baseline fluorescence of the acquisitions between trials, Takagaki et al. (2008) show that setting the gain as equal to the baseline fluorescence (typically estimated by averaging the first frames) induces a strong bias due to different bleaching kinetics between gain and baseline fluorescence. They propose to estimate the gain as the peak-to-peak amplitude of epileptiform spikes elicited by injection of bicuculline methiodide. Unfortunately this is not practicable in most physiological experiments. Moreover, in their setting, the way they measure the baseline fluorescence is unclear and could also induce bias.

2.3. Heartbeat Triggered Averaged

The heartbeat triggered average method takes advantage of the fact that an electrocardiogram is recorded simultaneously to the acquisitions and that all acquisitions are triggered at the same phase of the electrocardiogram. The method of EKG-locked predictors is introduced by Arieli et al. (1995b). Chen et al. (2008) model the heartbeat artifact as the repetition of a pattern approximated by the concatenation of two “half-Gaussian”: one for a rising edge and one for a falling edge. Width and amplitude of those Gaussians are fitted on all available blanks. The resulting heartbeat artifact of a given acquisition is reconstructed by concatenation of several copies of this pattern, shifted in time such that their centers correspond to the peaks of the simultaneously recorded ECG, accounting for possible variations of heartbeat durations with time. The amplitude of each pattern is fixed constant along heartbeats to avoid overfitting. In the subsequent analysis, the authors average the time courses within a spatial region of interest in order to obtain a single time course for each acquisition; so it seems that only one heartbeat time course by acquisition (*i.e.* as opposed to a heartbeat time course for each pixel) is reconstructed and then subtracted to the acquisition. Other heartbeat triggered average procedures for estimating and subtracting the heartbeat artifact can be found in the literature for ECG-locked VSDOI acquisitions, see for instance Arieli et al. (1995a); Ma et al. (2004); Lippert et al. (2007).

2.4. Automatic Component Separation Methods

Methods like heart triggered average or bleaching fit are based on *a priori* knowledge about the biophysical origin of those components. The approach is to reconstruct them according to this prior knowledge (e.g. heartbeat is synchronous with ECG), and subtract the result from the observation. Other attempts have been made to identify and retrieve all or parts of the components, directly from the acquisition with help of

statistical estimation methods. In particular, principal component analysis (PCA, Jolliffe (2002)) and independent component analysis (ICA, Hyvärinen and Oja (2000)) are reported to be quite efficient. Both methods rely on the hypothesis that the signal is decomposed as a sum of components (called *modes*) that are uncorrelated (PCA) or independent (ICA) across observations.

Maeda et al. (2001) and Inagaki et al. (2003) consider ICA on the temporal domain, within each single trial, to separate neuronal activity from heartbeat and respiration (no bleaching is described). Later, Reidl et al. (2007) propose an ICA on the spatial domain within each single trial. The authors normalize their data by zero-frame (see § 2.1) and then subtract a bleaching time course estimated as a low-pass filtered average of blanks. After ICA they obtain spatial modes, which enable to identify functional maps over the studied cortical surface. As a refined denoising process after blank subtraction, Onat et al. (2011a,b) use, on each single trial, *singular value decomposition* (SVD) of the space-time data matrix (*i.e.* the matrix obtained after concatenation of each frame reorganized as a column vector). They obtain several spatiotemporal modes defined as the separable outer product between the left singular vectors (which are nothing but the principal components of an uncentered spatial PCA) and the corresponding right singular vectors (which are the principal components of an uncentered temporal PCA). Then they classify the spatiotemporal modes according to the corresponding singular values, and to some other information such that the oscillatory behavior of the time course of the modes. Some are attributed to neuronal activity, some to artifacts, and the remaining ones (corresponding to small singular values) to residual noise.

In general, though ICA and PCA provide convenient ways for decomposing VSDOI signals, it is difficult to evaluate the nature (noise, artifact, or signal) of the resulting modes; if even possible. Indeed, the statistical hypotheses are strong (for instance PCA leads to orthogonal modes) and learning the modes on the data raises theoretical and practical difficulties. Usually the underlying statistical model is either on the spatial modes (in that case, each frame is an observation) or on the temporal modes (each pixel is an observation). A full spatiotemporal model would require an enormous amount of trials to be statistically accurate.

2.5. Linear Model

A last method relying on the decomposition into several modes has been developed in (Reynaud et al., 2011). The authors use the general linear model (GLM¹) framework, widely used in fMRI (Friston et al., 1995). It is fundamentally different from PCA or ICA. Indeed in this framework the modes (called regressors) must be known *prior* to the decomposition. The idea is to introduce a set of regressors for each component characterizing the response to the stimulus (or “signal”), and to find a linear combination of those regressors which best

approximates those components in the observed signal. The coefficients of the linear combination are taken to minimize the distance between the reconstructed and the observed signal. As long as the introduced regressors are all together linearly independent, the set of minimizing coefficients is unique and is the solution of a small size linear system. If the regressors representing the artifacts and those representing the neuronal activity are decorrelated enough then the resulting linear reconstruction should separate well the artifacts from the neuronal activity. Meanwhile if the regressors are well chosen the reconstruction should be accurate. One important problem is the fact that the temporal structures of several artifacts and responses components, such as heart beats and VSDOI responses particularly, have strong correlation and could not be considered as independent regressors in a linear model.

The crucial task in this approach is the design of the regressors. Each set of regressors constitutes a linear approximation basis, reducing the dimensionality of the component to the number of corresponding regressors. If this dimension is too high, the linear approximation of the component may overfit and capture noise and other components. If it is too low, the component cannot be recovered properly. In Reynaud et al. (2011) the authors take advantage of the strong temporal morphology of the components, refining their estimation through experiments designed to isolate as much as possible each component from the others. In brief, they identify three components on top of the neuronal response: bleaching, heartbeat and illumination fluctuation. Bleaching is modeled as a decaying exponential (*i.e.* with a regressor decaying exponentially at a given rate) and the two others are modeled as oscillating components (*i.e.* with sinusoidal regressors at a given set of frequencies). Fluctuation of the light source is first characterized on recordings of an illuminated paper surface, *i.e.* no dye is present. After a power spectrum analysis, five harmonics of a dominant fundamental frequency are selected. Then the bleaching time constant and the fundamental frequency of the heart are estimated simultaneously by fitting blank recordings. Two harmonics are introduced for the heart. Finally, a constant regressor is added to account for the baseline fluorescence.

In the GLM framework, reducing the dimensionality of the response space is the most delicate part. It strongly depends on the expected complexity of the response and on the information one wishes to extract from the VSDOI acquisition. Consider for instance the problem of establishing a map of orientation selectivity on the primary visual cortex of a mammal (as we do in § 4.2). A classical approach is to present full field, drifting contrast gratings of well-chosen spatiotemporal frequency. In response to such visual stimuli one expects VSDOI to record an overall synchronous raise of neuronal activity. The level of activity reached at a given pixel depends on the orientation selectivity of the underlying neurons, and the space resolution of VSDOI allows discriminating between several preferred orientation domains. In such a setting, the information one wishes to recover is only a scalar per pixel, namely the reached level of activity. Thus the response component’s dimensionality can be reduced to only one, by introducing a single “gate-shaped” response regressor that mimics roughly the temporal morphology

¹not to be confused with Generalized Linear Model, a broader class of models in statistics.

of the response: zero-valued up to some plausible delay after stimulus onset, then nonzero constant up to the stimulus offset, then back again to zero. In [Reynaud et al. \(2011\)](#) the stimuli used are also simple but more localized in space. Similar response morphology is expected but with delays varying from pixel to pixel. Hence the response space is also constituted by “gate-shaped” vectors, but with varying rising and decreasing times. SVD is further applied to the set of such vectors with all plausible delays to finally retain a ten-dimensional response space able to capture most of the considered spatiotemporal dynamic.

2.6. Towards Better Denoising Methods

In general, it is clear that VSDOI has not yet released all its potential for *in vivo* studies of the cerebral cortex. With the notable exception of 2-photon imaging, it is the only modality that provides real-time monitoring of neuronal activity at mesoscopic, population level; and understanding this scale of organization is a mandatory step for understanding brain function. During the last decades outstanding improvements have been made regarding the experimental technique, whether it is on the quality of the dyes, the building of the experimental setup, or the optical device. Even though developments of the hardware aspects of VSDOI, *i.e.* attempts to overcome the biological and physical limitations presented above, can still be expected, we believe that some efforts remain to be made on its software counterparts, *i.e.* better signal modeling and application of state-of-the-art signal processing techniques. We sum up here the general principles that, besides the previous approaches presented above, have driven our work.

2.6.1. Using the Spatiotemporal Structure

On top of their strong temporal morphologies, many components in the VSDOI signal have spatially structured origin, *e.g.* heartbeat artifact is stronger on veins, vibrations and illumination fluctuations affects the whole field, and neuronal response is organized in functional units or can exhibit wave-like patterns. In spite of the variety of VSDOI processing methods, very few seem to take advantage of the entire spatiotemporal structure of the observations. Consider for instance PCA or ICA on the spatial domain (§ 2.4). The time frames are modeled as independent observations, and permuting them randomly results in exactly the same spatial modes. Hence, temporal structure is not taken into account while learning the different components. Similarly but the other way around, the GLM proposed in [Reynaud et al. \(2011\)](#) (see § 2.5) processes each pixel independently, without taking into account any spatial structure. As for the SVD used in [Onat et al. \(2011a,b\)](#), the signal is decomposed into modes that are spatiotemporal but separable along space and time, limiting drastically the dynamic that can be retrieved from a few modes. Let us not that an extension of a temporal PCA is used in [Omer et al. \(2013\)](#) to account for the spacial structure of the signal.

Recently, [Yavuz \(2012\)](#) proposed a procedure for assessing and refining the source separation provided by the GLM method

of [Reynaud et al. \(2011\)](#), mixing both spatial and temporal information. First, a components separation in the temporal domain is obtained using GLM, and the result is divided into two groups of components: one comprising the sum of all artifacts (but the bleaching), and the other containing the neuronal response and the residuals. Then, spatial PCA is performed on each of those groups, and on the group of artifacts found within the blank acquisitions. Finally, the resulting components are compared with the blank components, based on the correlation between their corresponding (temporal) coefficients. Artifacts components presenting low correlation with the blank are re-labeled as neuronal activity components, and neuronal activity components presenting high correlation with the blank are re-labeled as artifacts. This approach has the double advantage of combining both model-based and statistical methods, and of using both temporal and spatial information. However, each step suffers from the aforementioned limitations of GLM and PCA, and combining temporal *then* spatial information in two separated steps might not be optimal.

2.6.2. Contributions

In general, separating the sources using spatiotemporal components is a difficult task. Even when the temporal dynamic of each pixel is simple, the relative delays and amplitudes from one pixel to another result in a wide variety of possible spatiotemporal patterns, and the dimensionality of the problem becomes prohibitively large. Capitalizing on recent advances in the field of inverse problem regularization in imaging, we propose to extend the GLM framework to include non-smooth convex regularizations. This corresponds to switching to a non-linear processing pipeline, which allows to include more complex data models that takes into account the transience and piecewise regular structures of the VSDOI signal. We call the resulting method “SMCS” (Spatially Structured Sparse Morphological Component Separation) since it makes use of ideas from the field of sparse regularization to perform the separation between the signal and the noise.

3. Methods

Within this section, we detail our model of the signal acquired by *in vivo* voltage-sensitive dye optical imaging and our theoretical approach for separating the neuronal activity from the artifacts. In order to fulfill the conclusive requirements of the previous section, our method relies on the *sparsity framework*.

In Section 2.5, we explained that in the GLM framework, the possible components that can be retrieved lie within a low-dimensional vector space, conditioned by their corresponding set of regressors. It is difficult to allow for more complex dynamics, since it requires to increase the dimensionality of those vector space. Indeed, the more correlated are the regressors of two different components, the less robust to noise is their separation capability. Increasing the complexity to the point where the set of regressors span the whole space results in a model which captures any artifact or noise. The solution of the source

separation problem is not unique as soon as regressors representing different components are not linearly independent; if two components have representation spaces with nonzero intersection, the signal along this intersection can be attributed arbitrarily to one component or to the other.

In the GLM terms, the idea of a *sparse* model is to introduce a large set of regressors, typically richer than within the GLM framework, modeling the signal, and seek for a sparse reconstruction where only a limited number of those regressors can be selected, *i.e.* most regressors must be weighted with coefficient zero.

This section contains important notations that will be used all along this paper, and the definition of the functionals used to enforce our assumptions on the signal. After setting the terms of our approach, we discuss the practical challenges that it raises before being applicable to voltage-sensitive dye optical imaging data.

3.1. Notations

We represent a VSDOI *acquisition* comprising T frames of P pixels as a matrix

$$\tilde{Y} \stackrel{\text{def}}{=} (\tilde{y}_{t,p})_{\substack{1 \leq t \leq T \\ 1 \leq p \leq P}} \in \mathbb{R}^{T \times P}.$$

The time course of the acquisition at the pixel indexed by $p \in \{1, \dots, P\}$ is the column vector $\tilde{Y}_p \stackrel{\text{def}}{=} (\tilde{y}_{t,p})_{1 \leq t \leq T} \in \mathbb{R}^T$.

The signal is decomposed into different spatiotemporal *components* $Y^{(c)} \stackrel{\text{def}}{=} (y_{t,p}^{(c)})_{t,p} \in \mathbb{R}^{T \times P}$ that have the same size as \tilde{Y} . The *residual* component, which is the part of the signal that cannot be explained by the previous components, is denoted $\tilde{R} \stackrel{\text{def}}{=} (\tilde{r}_{t,p})_{t,p} \in \mathbb{R}^{T \times P}$. This residual is modeled as a Gaussian white noise whose standard deviation varies in space and time; the matrix of those standard deviations is denoted $\tilde{\Sigma} \stackrel{\text{def}}{=} (\tilde{\sigma}_{t,p})_{t,p} \in \mathbb{R}^{T \times P}$. In a similar fashion, since the *gain* might vary in space and time, we store its values in a matrix $G \stackrel{\text{def}}{=} (g_{t,p})_{t,p} \in \mathbb{R}^{T \times P}$.

The time course of all pixels of a component (c) is linearly estimated in a certain *dictionary* of K_c *regressors* (also called *atoms*),

$$D^{(c)} \stackrel{\text{def}}{=} (d_{t,k}^{(c)})_{\substack{1 \leq t \leq T \\ 1 \leq k \leq K_c}} \in \mathbb{C}^{T \times K_c}.$$

The time course of the regressor indexed by a given $k \in \{1, \dots, K_c\}$ is then the column vector $D_k^{(c)} \stackrel{\text{def}}{=} (d_{t,k}^{(c)})_t \in \mathbb{C}^T$.

The *coefficients* of the linear representation constitute the matrix

$$X^{(c)} \stackrel{\text{def}}{=} (x_{k,p}^{(c)})_{\substack{1 \leq k \leq K_c \\ 1 \leq p \leq P}} \in \mathbb{C}^{K_c \times P}.$$

More precisely, we have for the pixel p of the frame t , $y_{t,p}^{(c)} = \sum_{k=1}^{K_c} d_{t,k}^{(c)} x_{k,p}^{(c)}$; in matrix notation $Y^{(c)} = D^{(c)} X^{(c)}$.

It is then possible to shorten further the notations by concatenating horizontally (respectively vertically) the dictionaries $D \stackrel{\text{def}}{=} (D^{(c)})_c \in \mathbb{C}^{T \times K}$ (respectively the coefficients matrices $X \stackrel{\text{def}}{=} (X^{(c)})_c \in \mathbb{C}^{K \times P}$), where $K \stackrel{\text{def}}{=} \sum_c K_c$. This way, we can write with matrix notation the sum of the components $\sum_c Y^{(c)} = DX$.

Some more matrix notations will be useful in the sequel. Given two matrices of the same size $M \stackrel{\text{def}}{=} (m_{i,j})_{i,j}$, $N \stackrel{\text{def}}{=} (n_{i,j})_{i,j} \in \mathbb{C}^{I \times J}$, we define the *termwise matrix multiplication* $M \cdot N \stackrel{\text{def}}{=} P \in \mathbb{C}^{I \times J}$ with $P \stackrel{\text{def}}{=} (m_{i,j} n_{i,j})_{i,j}$, and the *termwise matrix division* $M/N \stackrel{\text{def}}{=} Q \in \mathbb{C}^{I \times J}$ with $Q \stackrel{\text{def}}{=} (m_{i,j}/n_{i,j})_{i,j}$. Also, we define the ℓ_2 -norm of M as $\|M\|_2^2 \stackrel{\text{def}}{=} \sum_{i,j} |m_{i,j}|^2$. Finally, we denote the *transpose* of M by tM and its trace $\text{tra}(M)$.

In general, a variable with a tilde (\sim) denotes a variable representing raw data, as it is recorded during the acquisition. Alternatively, a variable with no tilde indicates that it has been normalized by the gain. For instance, we define $Y \stackrel{\text{def}}{=} \tilde{Y}/G$.

3.2. The Model

Using above notations, our model for the VSDOI signal is

$$\tilde{Y} = G \cdot (\sum_c Y^{(c)}) + \tilde{R}. \quad (1)$$

We now detail how each of these terms is modeled within our framework.

3.2.1. Gain and Residual

The spatiotemporal matrix G represents the gain between the amplitude of the components and the intensity of recorded fluorescence. The gain is supposed to vary only from one pixel to an other, and to be constant for the duration of an acquisition. That is, for each pixel p , there exists $g_p \in \mathbb{R}$ such that for all time t , $g_{t,p} = g_p$. The residual \tilde{R} takes into account modeling errors and random noise that corrupts the acquisition. We model mathematically this error as the sum of two independent random terms, one Poissonian term accounting for fluorescence and camera shot noise, and one Gaussian accounting for every other contributions. Now, a Poissonian variable with large mean can be seen as a sum of many smaller, independent Poissonian variables. By the central limit theorem, this is well approximated by a Gaussian variable whose standard deviation is proportional to its mean. As a result, the errors at each pixel of each frame are supposed to be realizations of Gaussians whose variances are determined by an affine function of the mean fluorescence intensity. During the sensory response, the variance covaries linearly with the mean. The intercept of this affine function determines a part of the noise which is constant across the whole acquisition, while its slope scales the influence of the shot noise. Recall that the baseline fluorescence dominates all other components, so that the level of fluorescence is essentially determined by the gain. As a consequence, noise level is also assumed to vary only from pixel to pixel and to remain constant along time, *i.e.* for each pixel p , there exists $\tilde{\sigma}_p \in \mathbb{R}$ such that for all time t , $\tilde{\sigma}_{t,p} = \tilde{\sigma}_p$.

3.2.2. Components Regressors

Following Reynaud et al. (2011), we consider three components, namely bleaching, periodic artifacts and neuronal activity, denoted respectively by (B), (P) and (R).

Bleaching. The regressors modeling the bleaching component are decreasing exponentials,

$$d_{t,k}^{(B) \text{ set}} = \exp(-t/\tau_k) , \quad (2)$$

for a set of time constants $\mathcal{T} \stackrel{\text{def}}{=} \{\tau_k\}_{1 \leq k \leq K_B}$. Since the bleaching is usually the component with lowest frequency and highest amplitude, we model the baseline fluorescence within the bleaching component by introducing one constant regressor, corresponding to $\tau_k = +\infty$. Moreover those regressors are highly correlated with the low-frequency regressors of the other components, making difficult their separation with the sole sparsity framework. Hence for this component the sparsity level is imposed, that is to say the dictionary $D^{(B)}$ is very restrained. In practice one or two regressors (on top of the constant one) are sufficient to capture most of the bleaching dynamics; introducing more regressors leads to overfitting. In consequence, the choice of the nonlinear parameter \mathcal{T} is crucial, see § 3.5.

Periodic Artifacts. The regressors modeling the periodic artifacts component are sinusoidal complex exponentials,

$$d_{t,k}^{(P) \text{ set}} = \exp(i2\pi f_k t) , \quad (3)$$

for a set of frequencies $\mathcal{F} \stackrel{\text{def}}{=} \{f_k\}_{1 \leq k \leq K_P}$. Since relevant frequencies are automatically selected thanks to sparsity regularization, this set can be as large as one wants. Nevertheless it is not worth introducing frequencies higher than the Nyquist-Shannon frequency (half the sampling frequency of the acquisition), and frequencies lower than the inverse of the acquisition duration are hardly detected. Moreover introducing too many frequencies increases vainly the computational cost for retrieving the coefficients of the model as proposed in § 3.2.3.

Neuronal Activity. Finally, as briefly motivated in the introduction, we model the neuronal activity with wavelet regressors. Wavelets are functions that are dilated, shifted replica of one given *mother wavelet* ψ , a compactly supported function with a certain level of regularity. Up to discretization, the wavelet regressors are thus defined as

$$d_{t,k}^{(A) \text{ set}} = \psi(2^{j_k} t - t_k) , \quad (4)$$

where t_k and j_k are respectively the *time* and *scale* of the wavelet indexed by k . More precisely, t_k ranges from 1 to T , and j_k ranges from 0 to $\log_2(T) - 1$, resulting in an order of $T \log_2(T)$ possible regressors. This allows to capture interesting features of a signal, smooth parts as well as transients and singularities, with a few regressors selected at the right temporal localization and scale.

Wavelet synthesis and analysis operators (multiplication by $D^{(A)}$ and its adjoint) are computed through filtering, successively applied along the scales (Mallat and Zhong, 1992). Low scale (*i.e.* low frequency) wavelets are often too correlated with other regressors to be useful in source separation. Hence, we usually compute only the steps for j ranging from a certain j_{\min} up to $\log_2(T) - 1$.

3.2.3. Retrieving the Components

In order to retrieve the coefficients of the model, our approach consists in two steps. First, the gain G and the noise level $\tilde{\Sigma}$ must be estimated (see § 3.5); we get a noisy estimate of the mixture of components by dividing the observation by the gain. We then retrieve the coefficients by solving an optimization problem,

$$\text{find } \hat{X} \text{ in } \underset{X \in \mathbb{C}^{K \times P}}{\text{argmin}} \frac{1}{2} \| (Y - DX) / \Sigma \|_2^2 + \Psi^{(\Lambda)}(X) . \quad (5)$$

The left-hand term in the summand is the data-fidelity term; the division by Σ accounts for the confidence one has on the observation depending on the noise level. Up to a constant, it is actually equal to the opposite of the log-likelihood of the sum of the components being DX given the noisy observation Y , under our statistical assumptions on the residual R (see § 3.2.1).

The right-hand term is a *regularization* ensuring that the resulting coefficients are relevant, in spite of the large number of regressors in D . Such approach, in which one defines an optimization problem whose solutions have the properties of the signal one is looking for, is called *variational*. In our model, Ψ is a sum of *penalizations* over the coefficients, notably promoting the sparsity of the solution: only a few coefficients can be nonzero. The next section describes the penalizations that we design for noisy source separation in VSDOI, and precise the dependancy on the parameters Λ .

3.3. Spatially Structured Penalizations

A classical sparsity promoting penalization is the ℓ_1 -norm, weighted by a nonnegative scaling parameter $\Lambda \stackrel{\text{def}}{=} (\lambda_{k,p})_{k,p} \in \mathbb{R}^{K \times P}$,

$$\|\Lambda \cdot X\|_1 = \sum_{k=1}^K \sum_{p=1}^P \lambda_{k,p} |x_{k,p}| . \quad (6)$$

However, this penalization considers each coefficients of each pixel individually; this is not robust enough to noise to provide good source separation for VSDOI. In order to benefit from the spatial structure of the components, we design some sparsity enforcing penalizations utilizing spatial information.

3.3.1. Periodic Artifacts: $\ell_{1,2}$ -Norm by Blocks

When performing VSDOI, fluctuations in the illumination supply affect the whole observation field; animal respiration and cortical movements affect entire spatial areas in unison; heartbeat artifacts are much stronger along the veins. In general, because of the biophysical origins of the periodic artifacts, when a given frequency affects the time course of one given pixel, chances are that the same frequency also affects neighbouring pixels, with similar intensity.

Thus, detection of significant frequencies is improved by replacing, in the ℓ_1 -norm, the modulus of a single coefficient by the norm of a group of coefficients of neighboring pixels. We call a *block structure* (over the k -th coefficient map) a family of *spatial blocks* of neighboring pixels. Then, given a family

$B^{(p)} \stackrel{\text{def}}{=} (B_k^{(p)})_{1 \leq k \leq K_p}$, where for each frequency k , $B_k^{(p)}$ is a block structure, we define the $\ell_{1,2}$ -norm by blocks

$$\|X^{(p)}\|_{\ell_1, \Lambda_B^{(p)}} \stackrel{\text{def}}{=} \sum_{k=1}^{K_p} \sum_{b \in B_k^{(p)}} \lambda_{k,b}^{(p)} \|X_{k,b}^{(p)}\|_2, \quad (7)$$

where for a block b , we denote $\|X_{k,b}^{(p)}\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{p \in b} |x_{k,p}^{(p)}|^2}$. The scaling parameter

$$\Lambda_B^{(p)} \stackrel{\text{def}}{=} \{B^{(p)}, (\lambda_{k,b}^{(p)})_{(k,b)}; 1 \leq k \leq K_p, b \in B_k^{(p)}\}$$

allows to weight differently the penalization on each block.

The choice of the block structures in $B^{(p)}$ is discussed in § 3.3.4. Note that in the particular case where for all $k \in \{1, \dots, K_p\}$, $B_k^{(p)}$ comprises all blocks restrained to individual pixels, i.e. $\forall k \in \{1, \dots, K_p\}$, $B_k^{(p)} \stackrel{\text{set}}{=} (\{p\})_{1 \leq p \leq P}$, then the $\ell_{1,2}$ -norm amounts to the ℓ_1 -norm (6).

To further strengthen the sparsity enforced by our model, we have used a reweighted $\ell_{1,2}$ -norm penalization, following an idea introduced in Candès et al. (2008). Details about this process can be found in Raguét (2014, Sec.II.3.2).

3.3.2. Neuronal Activity: $\ell_{1,2}$ -Norm for Sparse Temporal Wavelets ; Total Variation for Sparse Spatial Gradients

Similarly to the periodic artifacts components, neuronal activity components of neighboring pixels in VSDOI are often highly correlated. For instance, many publications on VSDOI report wave-like patterns of activity (Wu et al., 2008), where at each time frame, regions of homogeneous activity are distinguishable. In the same time, sharp differences of activity can be present between these regions.

Thus, we penalize the neuronal activity component by an other $\ell_{1,2}$ -norm, parameterized by

$$\Lambda_B^{(A)} \stackrel{\text{def}}{=} \{B^{(A)}, (\lambda_{k,b}^{(A,\ell)})_{(k,b)}; 1 \leq k \leq K_A, b \in B_k^{(A)}\}.$$

Note however that $\ell_{1,2}$ -norm is less adapted for neuronal activity than for the periodic artifacts component. Indeed, it tends to select groups of coefficients with similar modulus, but without regards to their signs. A given frequency may influence a spatial area with different phases from pixel to pixel. On the contrary, neuronal activity time course at pixels with similar intensity must have the same sign.

This can be modeled in our sparsity framework as the hypothesis that the spatial gradient of each spatial map of coefficients is sparse. The spatial gradient is defined by approximating the spatial derivatives in each direction by finite differences. More precisely, given a coefficient index $k \in \{1, \dots, K_A\}$ and a block b of size 2×2 pixels, the spatial gradient of the k -th coefficient map of $X^{(A)}$ at the center of b is defined as the four-dimensional vector²

$$\Delta X_{k,b}^{(A)} \stackrel{\text{def}}{=} (x_{k,p}^{(A)} - x_{k,b}^{(A)})_{p \in b} \quad (8)$$

²the notation Δ for the discrete spatial gradient emphasizes the fact that the spatial derivatives are approximated by finite differences; it should not be confused with the usual notation for the Laplacian operator.

where we denote $x_{k,b}^{(A)} \stackrel{\text{def}}{=} \frac{1}{4} \sum_{p \in b} x_{k,p}^{(A)}$ the average value of the coefficients within the block b . Note that other choices for the finite differences are possible.

Then, we enforce sparsity on the spatial gradients of the coefficients $X^{(A)}$ by using the two-dimensional total variation semi-norm penalization over each coefficient map,

$$\begin{aligned} \|X^{(A)}\|_{\delta, \Lambda_{TV}^{(A)}} &\stackrel{\text{def}}{=} \sum_{k=1}^{K_A} \sum_{b \in B_{TV}} \lambda_{k,b}^{(A,\delta)} \|\Delta X_{k,b}^{(A)}\|_2, \\ &= \sum_{k=1}^{K_A} \sum_{b \in B_{TV}} \lambda_{k,b}^{(A,\delta)} \sqrt{\sum_{p \in b} (x_{k,p}^{(A)} - x_{k,b}^{(A)})^2}, \end{aligned} \quad (9)$$

where B_{TV} is the block structure containing all possible blocks of size 2×2 , and

$$\Lambda_{TV}^{(A)} \stackrel{\text{def}}{=} \{B_{TV}, (\lambda_{k,b}^{(A,\delta)})_{(k,b)}; 1 \leq k \leq K_A, b \in B_{TV}\}$$

allows to weight differently the penalization over each block.

3.3.3. Some Suitable Hard Constraints

Using the variational framework that we defined in § 3.2.3 also allows to enforce other properties of the signal we are looking for, not necessarily related to sparsity. Here, we define penalizations which are useful for VSDOI component separations. They are dubbed “hard constraints” because they all have in common that they bear no compromise: they define subsets of the optimization space in which the solution of (5) must lie, whatever might be the observations or other constraints in the problem. They are defined as *convex indicator functions* which are functions that take value 0 within a certain set and $+\infty$ outside.

Bleaching: Nonnegative Orthant. The bleaching component captures the fluctuations of the baseline fluorescence, which is by nature a positive quantity. Since bleaching regressors are nonnegative, the bleaching coefficients are necessarily nonnegative as well. We ensure this with the convex indicator function of the *nonnegative orthant*

$$\iota_+(X^{(B)}) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \forall (k, p), x_{k,p}^{(B)} \in \mathbb{R}_+, \\ +\infty & \text{otherwise.} \end{cases} \quad (10)$$

Periodic Artifacts: Bounded Amplitude. In order to ensure stability of the component separation results and enforce prior knowledge over the components amplitudes, one can bound from above the coefficients of the periodic artifacts. We introduce a hard constraint which is spatially structured similarly to the $\ell_{1,2}$ -norm by blocks, as

$$\iota_{\ell_{M_B^{(p)}}}(X^{(p)}) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \forall (k, b) \|X_{k,b}^{(p)}\|_2 \leq \mu_{k,b}^{(p)}, \\ +\infty & \text{otherwise,} \end{cases} \quad (11)$$

where $M_B^{(p)} \stackrel{\text{def}}{=} \{B^{(p)}, (\mu_{k,b}^{(p)})_{(k,b)}; 1 \leq k \leq K_p, b \in B_k^{(p)}\}$

sets the maximal amplitude of each spatial block of coefficients.

Neuronal Activity: Bounded Deviation and Known Bounds. In a similar fashion, one can control the amplitude of the spatial gradients of the neuronal activity coefficients with a penalization which is spatially structured as the total variation, as

$$\iota_{\delta M_{TV}^{(A)}}(X^{(A)}) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \forall (k, b), \|X_{k,b}^{(A)}\|_2 \leq \mu_{k,b}^{(A)}, \\ +\infty & \text{otherwise,} \end{cases} \quad (12)$$

$$\text{where } M_{TV}^{(A)} \stackrel{\text{def}}{=} \{B_{TV}, (\mu_{k,b}^{(A)})_{(k,b)}; 1 \leq k \leq K_A, b \in B_{TV}\}$$

sets the maximal amplitude of each spatial gradients.

Moreover, it might happen that in some VSDOI experiments, the experimentalist knows in advance that at some points in time, no neuronal activity should be expected; or, on the contrary, that the neuronal activity should be above a certain value. This can be easily transposed in our variational framework, again with a convex indicator function

$$\iota_{[\underline{Y}^{(A)}, \bar{Y}^{(A)}]}(Y^{(A)}) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \forall (t, p), Y_{t,p}^{(A)} \in [\underline{Y}_{t,p}^{(A)}, \bar{Y}_{t,p}^{(A)}], \\ +\infty & \text{otherwise,} \end{cases} \quad (13)$$

where the spatiotemporal bounds $\underline{Y}^{(A)}$ and $\bar{Y}^{(A)}$ define respectively the minimum and maximum values of the neuronal activity. Some or all of those values can be $\pm\infty$ when no bound is known.

3.3.4. Structure of the Parameters

Altogether, the penalization $\Psi^{(\Lambda)}$ in (5), where Λ stands for the set of all parameters involved, expands for all $X \in \mathbb{C}^{K \times P}$, as

$$\begin{aligned} \Psi^{(\Lambda)}(X) \stackrel{\text{set}}{=} & \|X^{(A)}\|_{\ell, \Lambda_B^{(A)}} + \|X^{(A)}\|_{\delta, \Lambda_{TV}^{(A)}} + \|X^{(B)}\|_{\ell, \Lambda_B^{(B)}} + \iota_{\delta M_{TV}^{(A)}}(X^{(A)}) \\ & + \iota_{[\underline{Y}^{(A)}, \bar{Y}^{(A)}]}(D^{(A)} X^{(A)}) + \iota_+(X^{(B)}) + \iota_{\ell M_B^{(B)}}(X^{(B)}). \end{aligned} \quad (14)$$

In all generality, there can be one different weight for each spatial block of each coefficient map involved in (14). In practice, one should reduce as much as possible the number of degrees of freedom of the model to avoid overfitting and to be able to set relevant parameters automatically (see the discussion § 3.5). Our approach is to set, for each component (c) and each regressor $k \in \{1, \dots, K_c\}$, a unique parameter. Such parameter can be denoted $\lambda_k^{(c)}$, it is characteristic of the significance of regressor k of component (c) over the entire spatiotemporal dynamic. The final map of weights is set as a function of the spatially varying residual level σ_p , proportionally to $\lambda_k^{(c)}$. We refer to [Raguet \(2014, Chap.VI, Sec.1\)](#) for a detailed description of this parameterization.

Since we are working with blocks instead of isolated coefficients, we extend the previous principle to block structures. For each component (c) and each regressor $k \in \{1, \dots, K_c\}$, we chose first a *block side length* $s_k^{(c)}$. Then, the family of blocks $B_k^{(c)}$ contains all possible square blocks of $s_k^{(c)} \times s_k^{(c)}$ pixels. Finally, for each of those blocks, the corresponding penalization weight is set now as a function of $\|\Sigma_b\|_2$, where $\|\Sigma_b\|_2 = \sqrt{\sum_{p \in b} \sigma_p^2}$, proportionally to $\lambda_k^{(c)}$.

The setting of maximum amplitude values $M_B^{(c)}$ is slightly different, because they should only depend on the underlying

signal amplitude and not on the noise level. Proportionally to an overall scaling, $\mu_k^{(c)}$, the maximum amplitude over each block is then only a function of the block size.

3.4. Optimization Algorithm

The resolution of (3.2.3) using the regularization defined in (14) corresponds to the resolution of a non-smooth convex optimization problem. Special care is required when solving such large scale problem, and we propose to use so-called *proximal splitting methods*. They have proven successful to solve highly structured optimization problem that typically involves sparsity enforcing terms; see in particular the review of [Combettes and Pesquet \(2011\)](#) for a panorama of these proximal splitting schemes and of their applications in image processing. For our SMCS problem, we advocate the use of the Generalized Forward Backward (GFB) splitting scheme of [Raguet et al. \(2013\)](#), which enables us to makes use of the structure of the functional to be minimized. We refer to [Raguet \(2014, Chap.IV\)](#) for a detailed description of how to formulate (3.2.3) in order to be manageable by the GFB algorithm.

3.5. Parameter Selection

The SMCS method, which corresponds to computing a solution of (3.2.3), depends on many parameters, most notably; the gain G ; bleaching constants \mathcal{T} ; noise level $\tilde{\Sigma}$; sparsity blocks and weights for periodic artifacts and responses $(\Lambda_B^{(P)}, \Lambda_B^{(A)})$, total variation weights and bounds for the response $(\Lambda_{TV}^{(A)}, M_B^{(P)}, M_{TV}^{(A)})$. In order to achieve good separation performance, it is crucial to carefully select these parameters.

Parameter selection for inverse problem regularization is a difficult problem in general, and there is no definite answer to this question. A review of some classical methods can be found in [Raguet \(2014, Chap.V\)](#). In our setting of VSDOI processing, we assume that we have at our disposal, together with the acquisitions in stimulus condition, a set of blank acquisitions, in which the neuronal activity $Y^{(A)}$ is neglected in a first approximation. This assumption allows us to approximate the separation process by a series of simpler denoising task, for which relevant parameters can be easily deduced once the noise level is estimated. We refer to [Raguet \(2014, Chap.VI\)](#) for a detailed description of this parameter selection procedure.

The assumption that ongoing brain activity contains no significant neuronal activity might be questionable when studying spontaneous cortical activity. In practice, we found that the spontaneous activity we are dealing with in that case is so important (typically composed of travelling waves) that it cannot be confused with ongoing background noise. A quick screening allows to identify easily the blank trials presenting such spontaneous activity. Such blanks are *not* used for estimating parameters, but are good examples of neuronal activity that cannot be repeated identically.

An important feature of our method is that it provides an automatic adaptation of the local processing scale in both space and time. Indeed, the model allows for one different spatial block size for each regressor, and these block sizes are selected through an automatic parameter selection method. Regarding

the temporal domain, the regressors for the neuronal activity are wavelets, which are atoms localized in time. Sparse wavelet approximation is known to be similar to an adaptive approximation on dyadic blocks, see for instance [Mallat \(2008\)](#).

4. Results

We are now set for applying our noisy component separation method to voltage-sensitive dye optical imaging data. Due to the difficulty of the task and the lack of *ground truth*, *i.e.* prior knowledge of what the neuronal signal should look like through voltage-sensitive dye optical imaging, this work remains essentially exploratory.

4.1. Synthetic Data

This section investigate the power of separability of our SMCS method, in a controlled, noisy condition, and compare it with the GLM method presented in [§ 2.5](#).

4.1.1. Random Generative Model

We only briefly sketch here the synthetic data generation process, and we refer to [Raguet \(2014, Sec.VII.2\)](#) for a detailed description. For blank and stimulus condition acquisitions, we simulate $T \stackrel{\text{set}}{=} 256$ frames of size $P \stackrel{\text{set}}{=} 50 \times 50$ pixels, with a sampling frequency of 500Hz and spatial resolution of $50 \mu\text{m}$ per pixels, leading to an acquisition duration of 512ms and observation field of $2.5 \times 2.5 \text{ mm}^2$. We define a random generative model of synthetic data, so that the resulting separations can be quantified in terms of average and standard deviation. More precisely, we generate ten realizations of $\tilde{Y}^{(b)} = G^{(b)} \cdot (Y^{(b,b)} + Y^{(b,p)}) + \tilde{R}^{(b)}$ and of $\tilde{Y}^{(s)} = G^{(s)} \cdot (Y^{(s,b)} + Y^{(s,p)} + Y^{(s,A)}) + \tilde{R}^{(s)}$, where each $\tilde{Y}^{(c)}$ and \tilde{R} are random variables inspired by our real data (see [§ 4.3](#) and [§ 4.2](#)) and VSDOI literature, mimicking recordings of propagating waves in the cat's visual cortex:

- the gains ($G^{(b)}, G^{(s)}$) are spatially varying (depends on p) with a Gaussian profile ;
- the baseline fluorescence is constant and the bleaching $Y^{(b)}$ is a single exponential decay with an amplitude and a characteristic time which are spatial Gaussian white noise ;
- the periodic artifacts $Y^{(p)}$ are composed of cosines of with spatially varying amplitude and phases ;
- the neuronal activity $Y^{(A)}$ is deterministic and corresponds to a propagating wave simulated by solving a partial differential equation with a localized source term ;
- the random noise \tilde{R} is a spatially varying white noise generated in accordance with [§ 3.2.1](#) ;

The resulting components and acquisitions can be visualized on [Figure 1](#).

4.1.2. Separation Methods

Before analyzing separation performances, let us first briefly discuss the parameters used for each method (see [Raguet \(2014, Sec.VII.2\)](#) for more details).

GLM Designs (see [§ 2.5](#)). We generate the regressor for the GLM in an oracle manner, in order to favor as much as possible this method. In particular, the bleaching regressor have the characteristic time used to generate $Y^{(b)}$; the periodic regressors have the frequencies used to generate $Y^{(p)}$; the response regressors are defined as the first left singular vectors of the SVD decomposition of the matrix $Y^{(A)}$ (which is known to be the best basis of a given size for approximating linearly the response, in term of ℓ_2 -norm). We consider two setups: one with ten response regressors (GLM 10), and one with fifteen response regressors (GLM 15).

SMCS Designs. We used SCMS with one decreasing exponential for the bleaching component, and a generic set of frequencies for the periodic artifacts that are not optimized for the considered synthetic signals. We consider two setup: the first one (SMCS) there is no box constraint, while in the second one (SMCS box) uses box constraints enforcing a neuronal response which is non-negative everywhere, and null outside frames of interest.

4.1.3. Results

Similarity indicators. We quantify the quality of the component separation by comparing each retrieved component with its target component, over the whole spatiotemporal domain. We use both the *signal-to-noise ratio*, expressed in decibels as

$$\text{SNR}(\hat{Y}, Y) \stackrel{\text{def}}{=} 20 \log_{10} \left(\frac{\|\hat{Y}\|_2}{\|\hat{Y} - Y\|_2} \right), \quad (15)$$

and the *correlation coefficient*

$$\text{CC}(\hat{Y}, Y) \stackrel{\text{def}}{=} \frac{\text{tra}(\hat{Y}^T Y)}{\|\hat{Y}\|_2 \|Y\|_2}. \quad (16)$$

Mean and standard deviation across trials of those similarity indicators, over neuronal response and periodic artifacts components, are shown in [Figure 2](#).

Component separation and denoising with SMCS. We see from [Figure 1\(d\)](#) that our method approximately retrieves the neuronal response component, in spite of a highly perturbed environment and no prior knowledge on the signal activation times.

Comparison with GLM. As can be seen on [Figure 2\(a\)](#) [0, 512]ms, GLM (blue) outperforms SMCS (dark blue) for the retrieval of the neuronal response waveform over the whole acquisition time course. This is explained by the fact that the response regressors in GLM are designed to retrieve only phenomena that take place between 200 and 300 ms. On the contrary, our method seeks for neuronal events along the whole acquisition time course, hence is more sensitive to artifacts occurring outside those time frames.

In order to evaluate *sensitivity* of the method to neuronal events, we compute the similarity indicators (15) and (16) restricted over the time frames between 200 and 300 ms. Results [Figure 2\(a\)](#) [200, 300]ms shows better similarity indicator for SMCS, but not for GLM. This is further illustrated in [Figure 3](#).

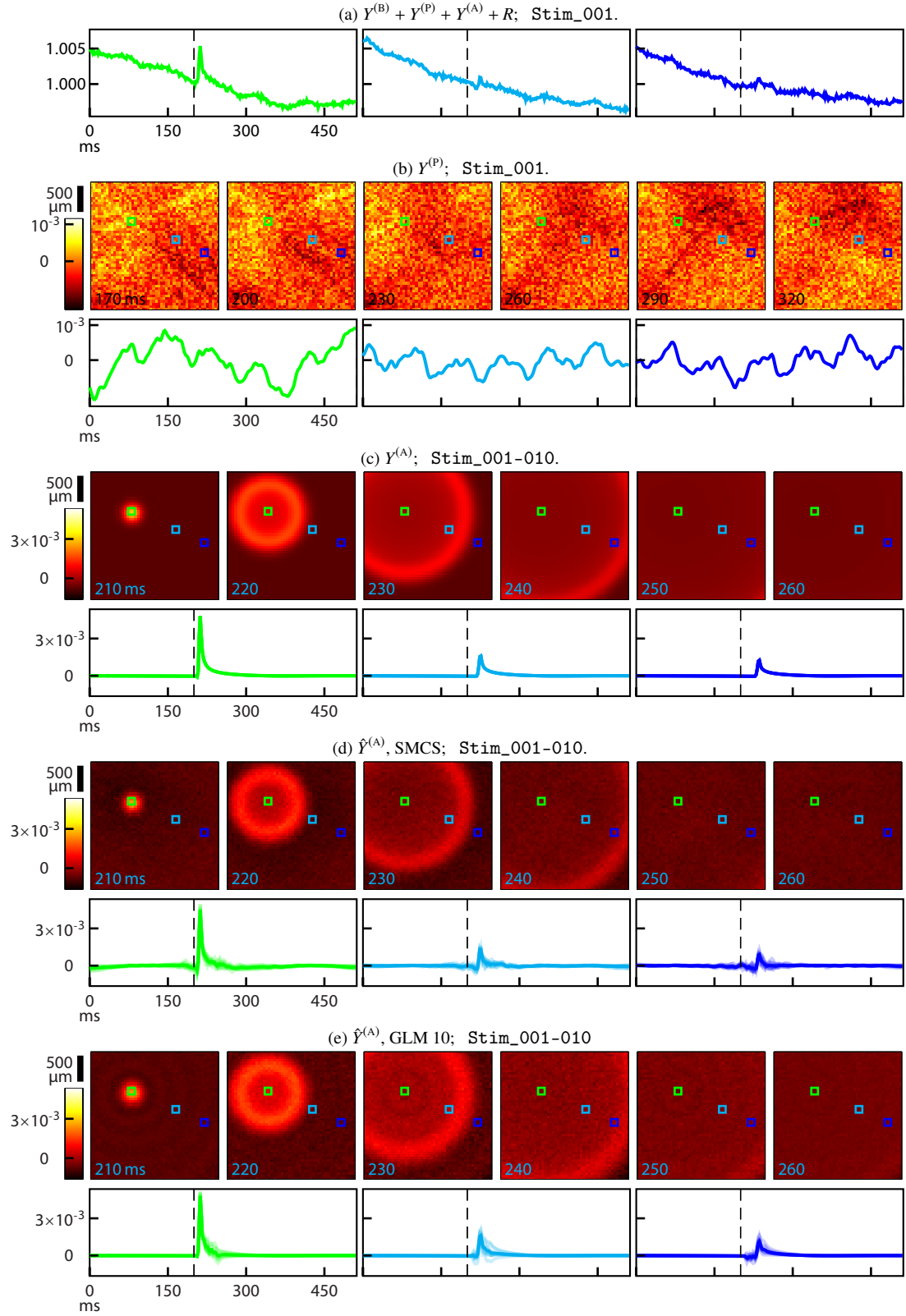


Figure 1: Synthetic data overview. Synthetic data and neuronal response recovery by SMCS and GLM 10. The one dimensional plotted time courses correspond to spatial positions indicated on the frames by a square of the same color. Each column represents one recording location in the cortical map. Vertical dashed line indicates stimulus onset. (a) Observations normalized by the gain, single trial. (b) Periodic artifacts, single trial. (c) Neuronal response, identical in all trials. (d) Recovery with SMCS, average across trials, single trials in transparency. (e) Recovery with GLM 10.

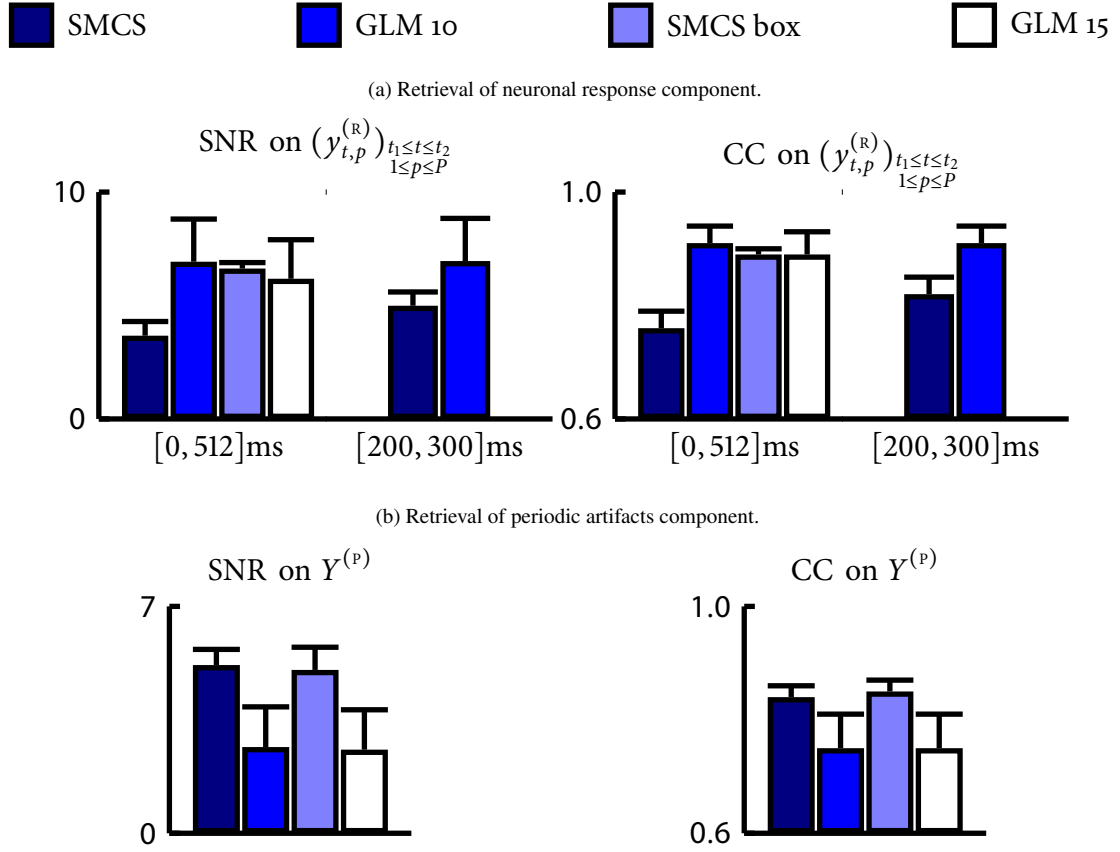


Figure 2: Performance of the component separation for SMCS and GLM methods on synthetic experiments. Results are given as average and standard deviation over ten realizations of the random generative model (see § 4.1.1). When specified, the signal-to-noise ratios (15) and the correlation coefficients (16) are computed over the frames included in $[t_1, t_2]$.

Then, *specificity* of the SMCS method to neuronal events is enhanced by introduction of the box constraints for SMCS box. We see on Figure 2(a) that SMCS box (light blue) performs on average as well as GLM 10. Moreover, it is more reliable than GLM since the latter exhibits higher standard deviation. This indicates that the separation performance of GLM on a given trial might depend on the realization of the periodic artifacts.

In fact, it turns out that good performances of GLM are explained by the accurate design of the response regressors (as explained above, § 4.1.2), and not by its component separation ability. To see this, we compute the similarity indicators for the retrieval of the periodic artifacts component. On Figure 2(b), it is obvious that periodic artifacts are better captured by both versions of SMCS (dark and light blue) than by GLM. Moreover, one might think that increasing the dimensionality of the response space in the GLM allows for better reconstruction accuracy. The white bars on Figure 2 show that this actually degrades its performances, because the response components gets mixed with noise and periodic artifacts; this can also be seen on Figure 3.

Finally, let us recall that according to the GLM approach proposed by Reynaud et al. (2011), the residual is considered to be part of the signal of interest, and should be added to the response component after component separation. This permits the GLM method to capture more complexity and trial-to-trial variability in the neuronal response, in spite of the restriction imposed over the response space. However, doing so in the current study makes no sense, since the performance indicators measure proximity to a deterministic response, known in advance, and from which the regressors are directly designed. Given the poor periodic artifacts reconstruction with GLM (Figure 2(b)), adding the residual would make the similarity indicators over the neuronal response component drop drastically.

4.2. Biological Data: Visual cortical Maps (Cat V1)

One of the most striking functional feature of most mammal's primary visual cortex is the *orientation selectivity*, first described by Hubel and Wiesel (1959). Many neurons in this cortical area are sensitive to elongated luminance contrasts that constitute edges in the visual field. Moreover, the intensity of the neuronal response depends on the similarity between the orientation of the edge within the visual field, and an orientation of reference, characteristic of the observed neuron. The closer to this *preferred orientation*, the higher the response.

Following an organization which is common to many cortical functional features, neurons sharing the same preferred orientation are gathered in *columns*, which are functional units extending spatially perpendicularly to the cortical surface (Mountcastle, 1997). In addition, in the cat's visual cortex, the variation of preferred orientation tangentially to the cortical surface is smooth (Hubel and Wiesel, 1962) (except at singularity points, where all the orientations are represented, see § 4.2.2). Thus, using optical imaging with sufficient spatial resolution, it is possible to identify the *dominant preferred orientation* underlying each recorded location (see for instance Bonhoeffer and Grinvald (1993)).

When presenting *full field, drifting luminance gratings* to the subject, a rise of activity is triggered in the entire primary visual cortex, which can last for the duration of the stimulus. Though most neurons are responding, their responses are modulated by orientation selectivity according to the orientation of the grating. When investigating time-independent orientation selectivity, the quantity of interest is only the *response intensity* so that there is no need for high temporal resolution for recording it. Thus, VSDOI is not the first choice experimental method for that purpose, and for recordings at the population level, less noisy methods like intrinsic optical imaging should be favored. Since our SMCS method aims at recovering complex temporal dynamics, it is precisely *not* designed for such experiments. However, orientation selectivity is one of the most understood functional properties of the primary visual cortex, and one prerequisite for delving into more complex features is to check that our denoising method does not suppress this information.

For that reason, we study in this section the use of the SMCS method for estimating orientation selectivity at the population level. We compare our method to the classical blank subtraction (§ 2.1) and to the GLM developed by Yavuz (2012).

It is important to note that the comparison of the orientation maps does not aim at establishing the superiority of SMCS, but rather at checking the SMCS does not miss some basic functional properties that simpler methods can retrieve. We would also like to stress that we do not study here functional maps such as dynamic single condition maps. A first reason is that the computation of these maps actually involves several trials (either by computing "differential maps" or a "cocktail blank"), while our aim here is to strictly study the behavior of the signal obtained after only one recording. A second reason is that the quality of the studied data does not seem to be good enough for finer investigation, such as the computation of dynamic orientation maps.

4.2.1. Data and Denoising Methods

The data set Cat_Gratings consists in 1280 ms long recordings of the primary visual cortex of a single cat, in area 18. The visual stimuli used are full field luminance gratings of four possible orientations, drifting in two possible directions, with a region of interest (ROI) limited to area 18. Each stimulus condition is recorded for ten trials, together with ten corresponding blank conditions.

Before analyzing the orientation selectivity, let us review the different component separation and denoising methods for extracting neuronal responses that we compare. Some resulting components for a given stimulus condition are shown for each method in Figure 4.

Blank Subtraction (BkS). All acquisitions of the protocol are triggered at the same phase of the simultaneously recorded ECG, and respiration of the subject is artificially controlled. Thus, the blank subtraction method introduced in § 2.1 is relatively efficient. We detail here our implementation.

Each trial under stimulus condition $\tilde{Y}^{(s)}$ is associated to its corresponding trial under blank condition $\tilde{Y}^{(b)}$. The gain of each

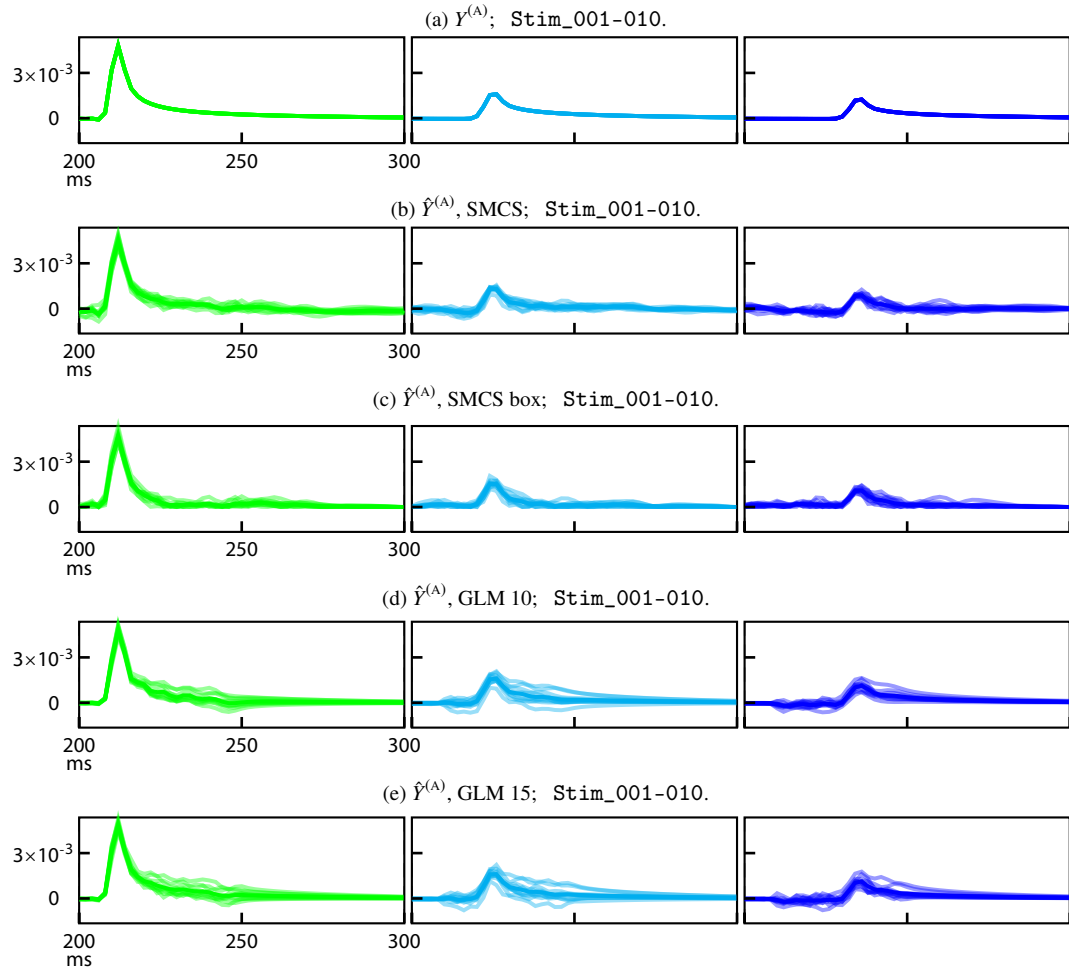


Figure 3: Close-up on the time courses of the targeted and retrieved synthetic neuronal responses. The plotted time courses correspond to spatial positions indicated on the frames of Figure 1 by a square of the same color. Results are shown as average across trials, together with the single trials in transparency.

acquisition is evaluated at each pixel p as the average fluorescence value over the frames before stimulus onset,

$$g_p^{(s)} = \frac{1}{t_{\text{on}}} \sum_{1 \leq t \leq t_{\text{on}}} \tilde{Y}_{t,p}^{(s)} \quad \text{and} \quad g_p^{(b)} = \frac{1}{t_{\text{on}}} \sum_{1 \leq t \leq t_{\text{on}}} \tilde{Y}_{t,p}^{(b)}.$$

Then, each acquisition is normalized by its gain, $Y^{(s)} = \tilde{Y}^{(s)}/G^{(s)}$ and $Y^{(b)} = \tilde{Y}^{(b)}/G^{(b)}$, and we create a smoothed blank acquisition $\Upsilon^{(b)} \stackrel{\text{def}}{=} (\nu_{t,p}^{(b)})_{t,p}$, getting rid of the high frequency noise by low-pass filtering with a Gaussian kernel. For all pixel p ,

$$\nu_{t,p}^{(b)} = \frac{\sum_{t' \in \{-w_f, \dots, w_f\}} \exp\left(-\frac{t'^2}{2\sigma_f^2}\right) y_{t+t',p}^{(b)}}{\sum_{t' \in \{-w_f, \dots, w_f\}} \exp\left(-\frac{t'^2}{2\sigma_f^2}\right)},$$

with symmetric boundary convention, *i.e.* $y_{t+t',p}^{(b)} \stackrel{\text{def}}{=} y_{1-(t+t'),p}^{(b)}$ if $t+t' < 1$, and $y_{t+t',p}^{(b)} \stackrel{\text{def}}{=} y_{T-(t+t'),p}^{(b)}$ if $t+t' > T$. In our experiment, we use $w_f \stackrel{\text{set}}{=} 10$ ms and $\sigma_f \stackrel{\text{set}}{=} 10$ ms. Finally, the neuronal response signal is estimated as $Y^{(s,A)} = Y^{(s)} - \Upsilon^{(b)}$.

GLM Design. We use the method described in Yavuz (2012, §6.2.2), which is a GLM approach (see § 2.5) inspired by the work of Reynaud et al. (2011), adapted to VSDOI recordings of neuronal responses to full field drifting luminance gratings in the cat's visual cortex. In this setting, there is *only one* response regressor, modeling rise and decrease of neuronal activity following respectively stimulus onset and stimulus offset. After source separation between neuronal signal, various artifacts, and residual, the residual is added to the response component, as advocated in both above references.

In contrast to our model (1) that integrates a gain matrix G (see Section 3.2.1 for more details about how it is estimated), in the work of Yavuz (2012), the choice was made *not* to model the gain. Keep in mind however that in the current study, the only quantity of interest is the polar angle of the vectorial sums (19) described below (§ 4.2.2). Now, even if the gain varies greatly from pixel to pixel, it does not vary much from trial to trial; hence, normalizing by the gain greatly affects the amplitude, but not the polar angle (20) of those vectorial sums.

The subsequent statistical refinement by PCA in the denoising proposed in Yavuz (2012) was not used here.

SMCS Design. We use the same parameters in this setting as in our synthetic data setting, § 4.1.2, except for $K_b \stackrel{\text{set}}{=} 3$, *i.e.* we model the bleaching with two decreasing exponentials (see § 3.2.2); we find $\mathcal{T} = \{123 \text{ ms}, 584 \text{ ms}\}$, and for periodic artifacts frequencies lower than 2 Hz. Since the acquisition duration is more than one second here, frequencies down to 1 Hz would be considered in our procedure; we discard them (*i.e.* they are not included in the set \mathcal{F} of Fourier regressors (2)) because the expected neuronal signal is much more important than respiration artifacts around those frequencies.

Qualitative Comparisons. As expected, the BkS method (Figure 4(a)), gives a very noisy signal. Still, averages across ten trials are informative about the simple, reproducible response to a

drifting grating. In comparison, time courses obtained by GLM are less noisy, even though the latter is added to the former, to obtain the final response, as explained above § 4.2.1. On Figure 4(c), the overall response dynamic is more prominent, relatively to the high frequency noise amplitude. Finally, traces obtained by SMCS are almost free of high frequency noise, while still exhibiting significant amplitude of trial-to-trial variability (Figure 4(e)).

Because the heartbeat and respiration are synchronized along trials, their retrieved average contribution presents distinct spatiotemporal patterns. On Figures 4(b) and (d), top rows, one can see fluorescence oscillations on large domains, with phases varying from left to right. Also, higher amplitudes distinguishes veins over the field of view. Note that because the considered GLM implementation does not model the gain (see § 4.2.1), it is not possible to compare directly its results to SMCS. However, periodic artifacts and neuronal responses have been normalized within each methods, in terms of color map and traces amplitude. Hence, we can see that amplitudes of retrieved periodic artifacts are less important in GLM than in SMCS, relatively to the corresponding neuronal responses. Moreover, periodic artifacts retrieved by SMCS exhibit more trial-to-trial variability, and more temporal complexity. We explain those facts by the richer dictionary of sinusoidal regressors modeling periodic artifacts within the SMCS, so that phase and frequency variations, both *inter*- and *intra*-trials, can be better captured. See in particular the significant contribution of the periodic artifacts around 12 Hz, Figures 4(d), middle trace³. We believe it to be an harmonic of the heartbeat artifact, since the corresponding pixel (in cyan color) is clearly situated on a vein.

However, higher sensitivity of the periodic artifacts component also presents some drawbacks. According to the SMCS results for the same cyan pixel, the neuronal activity averaged across trials starts to decrease *before* stimulus offset. Given the relative brevity of the stimulus duration, this is not expected; indeed, both BkS and GLM results show a longer sustained neuronal activity at this pixel. We explain the error made by SMCS by the fact that around stimulus offset, the heartbeat artifact at this location is at a rising phase, as indicated by brighter colors along the vein, on the frame 800 ms of both Figures 4(b) and (d).

An other significant difference between GLM and SMCS is the important decrease of the neuronal response component, occurring around 1 s and until the end of the acquisition. We believe it to be due to intrinsic artifact (described in § 1.3.2), because of its slow dynamic and emergence after significant neuronal response. The same trend can be seen on the BkS traces, but much fainter. We believe that this effect is overestimated by the SMCS method because the constant regressor (within the bleaching dictionary, see § 3.2.2) is not penalized and tends

³note that in our data, 12 Hz signal is absent from the periodic artifacts retrieved by GLM because this frequency is actually not included in the regressor basis. However, as stated by Yavuz (2012, 6.2.3), this artifact is still badly captured by GLM when introducing it. The reason for not including this frequency *at all* is that subsequent statistical refinements proposed by this author give better results that way.

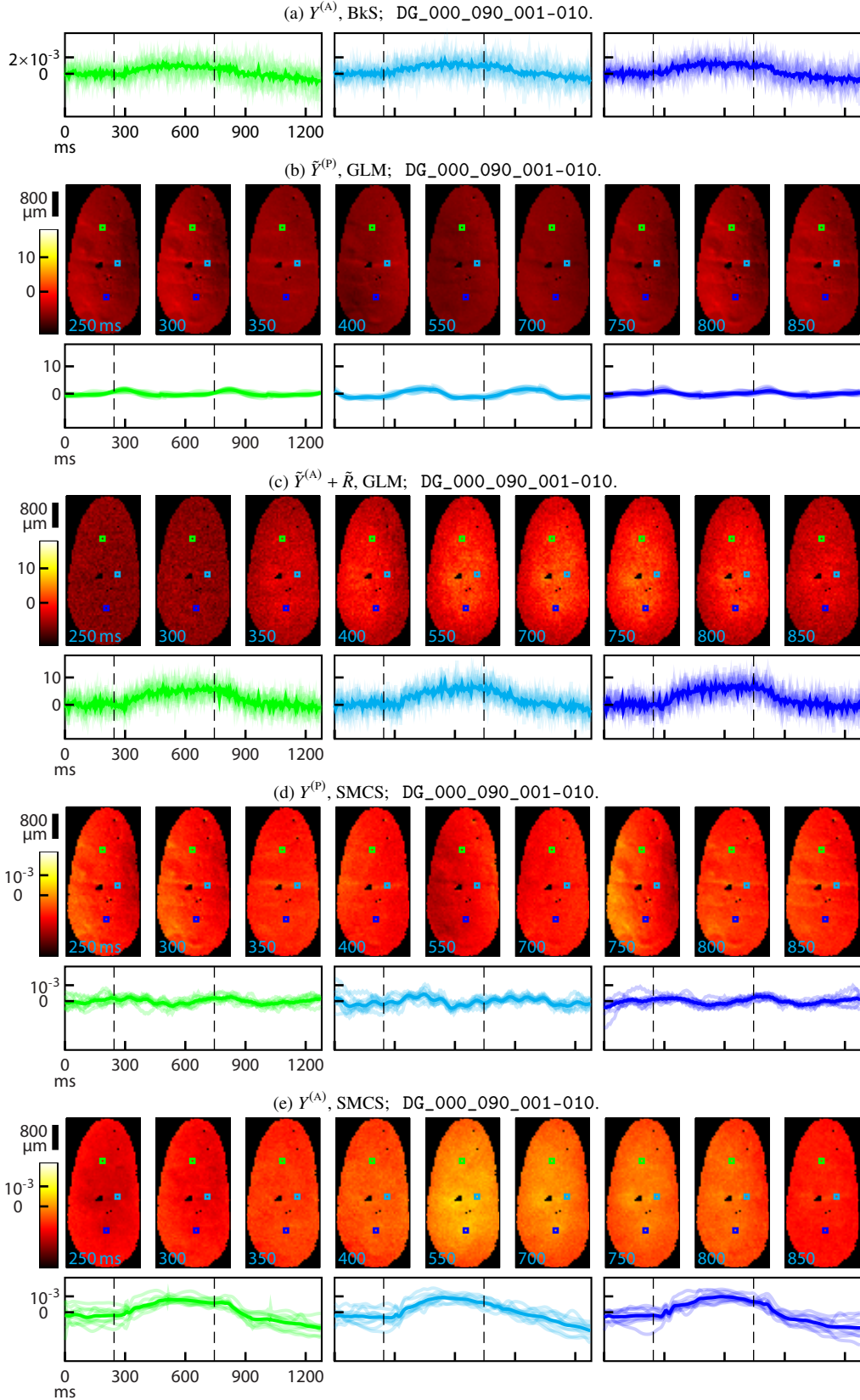


Figure 4: Component separations for the protocol Cat_Gratings. Component separations for the protocol Cat_Gratings, stimulus condition: full field, gratings of orientation 0° (horizontal), drifting direction 90° (upwards). The time courses correspond to spatial positions indicated on the frames by a square of the same color. Vertical dashed line indicates stimulus onset and offset. (a) Neuronal responses retrieved by BkS. (b) Periodic artifacts retrieved by GLM. (c) Neuronal responses retrieved by GLM. (d) Periodic artifacts retrieved by SMCS. (e) Neuronal responses retrieved by SMCS. Results are shown as average across trials, together with single trials time courses plotted in transparency. Note that GLM results are given in arbitrary units of recorded fluorescence level, since they have not been normalized by the gain (see § 4.2.1).

to adjust, at each pixel, to the mean value of the signal along time. This would also explain the slightly negative values before stimulus onset, observed on the mean traces of Figure 4(e).

Finally, let us note the ability of our SMCS method to capture, at the single trial level, a transient phenomenon that takes place between 40 and 80 ms after stimulus onset (*i.e.* around 300 ms after beginning of acquisition). After first increase, evoked activity slightly and rapidly decreases, before increasing again. This phenomenon, coined *deceleration-acceleration notch*, have been documented in the literature, see in particular Sharon and Grinvald (2002).

4.2.2. Estimation of Preferred Orientation

A VSDOI protocol investigating orientation selectivity over the whole cortical sample should usually contain recordings of neuronal responses to a set of full field drifting luminance gratings of various orientations, taken in a set $\{\theta_1, \dots, \theta_N\}$, sampling uniformly all possible orientations from 0° to 180° (in the case of our data set, $N \equiv 4$). We describe here a procedure for extracting static (*i.e.* we do not study evolution along time) orientation selectivity from such neuronal responses, $(Y^{(s,A)})_s$ (which can be extracted by any denoising method).

Mean Response Level. In order to estimate the dominant preferred orientation underlying a given spatial location in VSDOI recordings, one must first extract the *mean response level* of that spatial location to each presented orientation of the protocol. We denote it $\tilde{Y}^{(m,A)} \stackrel{\text{def}}{=} (\tilde{y}_{n,p}^{(m,A)})_{\substack{1 \leq n \leq N \\ 1 \leq p \leq P}}$, computed for each pixel p and orientation θ_n as

$$\tilde{y}_{n,p}^{(m,A)} = \frac{1}{|s(\theta_n)|} \sum_{s \in s(\theta_n)} \sum_{t_1 \leq t \leq t_2} y_{t,p}^{(s,A)}, \quad (17)$$

where $s(\theta_n)$ denotes the set of all trials of all grating stimuli with orientation θ_n within the protocol, and $t_1 \leq t_2$ delimit the time frames where the most significant response is expected.

In our numerical application, we use $t_1 \stackrel{\text{set}}{=} t_{\text{on}} + 100$ ms and $t_2 \stackrel{\text{set}}{=} t_{\text{on}} + 500$ ms = t_{off} .

Normalization Step. On our data set, using the mean response levels directly as defined in (17) for deducing preferred orientation is not satisfying. As can be seen on Figure 5(a), this leads to an overwhelming representation of some preferred orientation angles; such a bias in the distribution is not expected, considering the size of our cortical sample and the distribution of preferred orientations generally accepted (see for instance Mountcastle (1997)). We alleviate this bias by dividing the mean response levels of each orientation by its average across pixels, obtaining $Y^{(m,A)} \stackrel{\text{def}}{=} (y_{n,p}^{(m,A)})_{n,p}$ such that for all p and n ,

$$y_{n,p}^{(m,A)} = \frac{P}{\sum_{1 \leq p' \leq P} \tilde{y}_{n,p'}^{(m,A)}} \tilde{y}_{n,p}^{(m,A)}. \quad (18)$$

This phenomenon might be explained by lateral exciting connections between functional columns enhancing drastically even a small imbalance in the distribution, but we leave this

point to future investigations. In the literature, the closest mention to such preprocessing that we could find is Sharon and Grinvald (2002, Note 24.), where the authors perform a spatial high-pass filtering of their data, in order “to remove nonreproducible global differences in the responses to different orientations”. However, besides the lack of clarity of this explanation, the *nonreproducible* terminology suggests that these differences are stimulus independent, in contrast to what we observe in our data set. The same need for spatial high-pass filtering is also reported in Chavane et al. (2011, Appendix), in order “to remove slow gradients”.

Orientation Maps. Finally, we compute the preferred orientation at each spatial location using the *vectorial sum* technique, as briefly described in Blasdel and Salama (1986). It represents the orientation selectivity with a two-dimensional vector, which is the sum of unit vectors of polar angle determined by a given orientation, weighted by the mean response level to that orientation. More precisely, denoting for an angle θ , $\vec{u}(\theta) \stackrel{\text{def}}{=} (\cos(\theta), \sin(\theta))$, the vectorial sum at pixel p is

$$\vec{v}_p = \sum_{1 \leq n \leq N} y_{n,p}^{(m,A)} \vec{u}(2\theta_n). \quad (19)$$

Note the factor 2 applied to each orientation, so that the range of orientations $[0, 180]^\circ$ is mapped to the complete circle $[0, 360]^\circ$ and that the mean response level to two orientations that are orthogonal cancel out in the sum. The preferred orientation angle at pixel p is then half of the polar angle of \vec{v}_p ; expressed in degrees,

$$\vartheta_p = \begin{cases} \frac{1}{2} \arccos\left(\frac{v_{p,1}}{|\vec{v}_p|}\right) & \text{if } v_{p,2} \geq 0, \\ 180 - \frac{1}{2} \arccos\left(\frac{v_{p,1}}{|\vec{v}_p|}\right) & \text{if } v_{p,2} < 0, \end{cases} \quad (20)$$

where we denoted the coordinates of the vectorial sum $\vec{v}_p \stackrel{\text{def}}{=} (v_{p,1}, v_{p,2})$ and its amplitude $\|\vec{v}_p\| \stackrel{\text{def}}{=} \sqrt{v_{p,1}^2 + v_{p,2}^2}$. This procedure provides a spatial map of preferred orientations angle $\Theta \stackrel{\text{def}}{=} (\vartheta_p)_{1 \leq p \leq P}$. The amplitude carries some interesting information about the overall response level and sharpness of orientation selectivity, but we do not use it in the current study.

Results. Resulting orientation angle maps are displayed on figure Figure 5. Observe first the apparent need for the normalization described in (18): on all three panels of Figure 5(a), preferred orientations around 60° are largely dominant; in contrast, the distribution of preferred orientations on Figure 5(b) is well balanced. This phenomenon appears very similar for both GLM and SMCS methods, and seems less pronounced with BkS denoising; we do not know the reason for that.

Now, after applying the normalization, the resulting orientation maps are in accordance with the literature. In particular (see Bonhoeffer and Grinvald (1993)), we observe a smooth variation of the preferred orientation along the cortical surface, covering the whole orientation space within a characteristic distance of approximately 1 mm, and organized around *pinwheels*, which are locations at the confluence of homogeneous area, around which all orientations are represented.

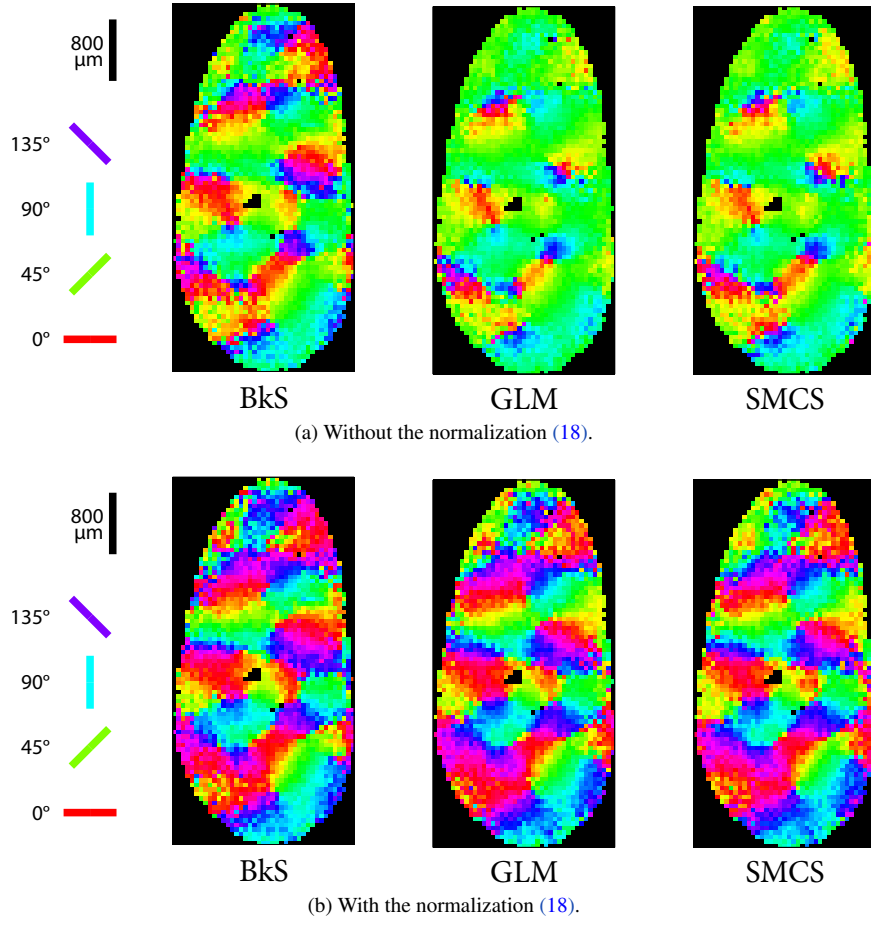


Figure 5: Resulting orientation angle maps Θ , computed with vectorial sum technique according to (19) and (20). The preferred orientation at each pixel is color coded in a circular color map; the colors corresponding to the four orientations of the stimuli used in the protocol Cat_Gratings are represented on the left side.

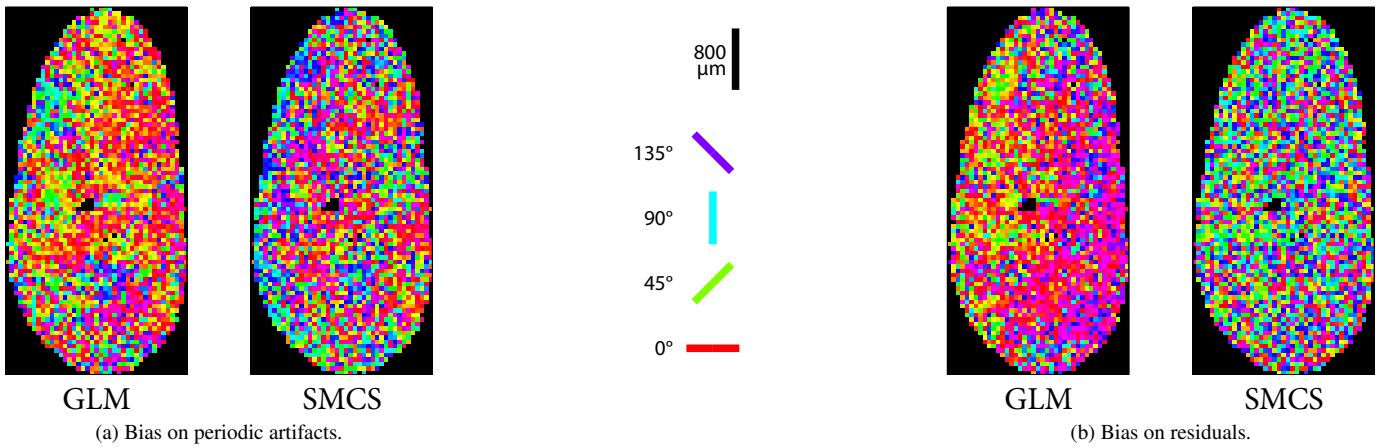


Figure 6: Orientation bias due to the presence of stimulus-related neuronal signal within non neuronal components, for the separations provided by GLM and SMCS methods. Those maps are computed as orientation angle maps but with non neuronal components. Note that normalization (18) makes no sense here, so it is not applied.

The three compared denoising methods lead to almost undistinguishable orientation angle maps. This suggests that, in spite of a strong denoising and very few priors over the neuronal signal, SMCS succeeds in capturing orientation selective information. In order to test further this fact, we then check for the presence of stimulus-related signals within other components of the separation. To do so, we apply exactly the same computations of preferred orientations as above, but using the periodic artifacts or the residual instead of the neuronal response. In the hypothesis that no stimulus-related signals are present, resulting maps should be meaningless, with random distribution of preferred orientations. On [Figure 6\(a\)](#), areas with dominant preferred orientations are visible, at the same spatial scale as the homogeneous areas in the orientation maps of [Figure 5](#). This proves that stimulus-related neural signal has been captured by the periodic artifacts during the process of component separation. Such bias seems weaker for the SMCS method, probably thanks to the hard constraints on the amplitudes of periodic artifacts (recall [§ 3.3.3](#)), which prohibit stronger periodic artifacts in acquisitions with stimulus conditions than in blank acquisitions. Now, the orientation bias on the residuals found by SMCS, [Figure 6\(b\)](#), appears to be completely random. Keep in mind that there still could be neuronal signal left out in the residual, like spontaneous activity or even evoked activity over short periods of time, but this is an encouraging result for our method.

In contrast, the residual found by GLM is strongly biased by the stimulus condition; this was however already expected, and constitutes an additional argument for merging the residual with the response component when using a GLM approach.

4.3. Biological Data: Somatosensory Cortex (Mouse S1)

In a last series of experiments, we explore visually the application of our method to VSDOI recordings of the mouse’s somatosensory cortex. Observation of propagating phenomenon have been reported in both awake and anesthetized animals, in response to direct sensory inputs as well as in ongoing neuronal activity, see in particular [Ferezou et al. \(2006\)](#).

The data set Mouse_A11 consists in 1024 ms long recordings of the somatosensory cortex of a single mouse, in the barrel cortex. In the stimulus condition, all the animal’s whiskers are briefly and simultaneously deflected, evoking a large raise of neuronal activity, first localized and then spreading rapidly to the entire recorded area ([Figure 7\(b\)](#)). We apply our method to nine recordings with stimulus condition. Moreover, out of twenty blank acquisitions considered, nine presented strong patterns of spontaneous activity, closely resembling the evoked propagations.

The neuronal response in those data is very large, and its overall shape is distinguishable after simple normalization of the raw data by an estimation of the gain ([Figure 7\(a\)](#) or [Figure 8\(a\)](#)). However, the level of high frequency noise is also important, and periodic artifacts are clearly visible. We apply the SMCS methods with the same parameters as in our synthetic data setting, [§ 4.1.2](#), except for $K_b^{\text{set}} = 3$, *i.e.* we model the bleaching with two decreasing exponentials (we find $\mathcal{T} = \{245 \text{ ms}, 5576 \text{ ms}\}$).

4.3.1. Stimulus Condition Acquisitions

We show on [Figure 7\(b\)-\(d\)](#) the denoising of a single trial, at some spatial positions and time frames of interest. We see that the high frequency noise has been well removed, except perhaps right after stimulus onset ($t_{\text{on}}^{\text{set}} = 200 \text{ ms}$, see that the amplitudes of R are smaller on frames 210 ms and 240 ms). Indeed, the activity here is so reproducible that the wavelets capturing the neuronal events occurring over this short time window are not penalized.

Moreover, a substantial part of the periodic artifacts seems to have been removed as well. Consider in particular the green pixel on [Figure 7\(c\)](#), whose time course is affected by consequent heartbeat artifacts, because it is located right over a vein crossing the field of view. However, remaining influence of the veins is still significant on [\(b\)](#), see in particular the first frame, 70 ms: the neuronal response component has captured periodic artifacts. This is in fact impossible to avoid completely, since we are seeking for *any* possible spatiotemporal neuronal activity, on data that are *not* synchronized with the heartbeat of the animal.

More worrisome, the confusion happens also the other way around. Indeed, each temporal trace of [\(c\)](#) is biased upwards just after stimulus onset; and the overall amplitude is higher on frame 240 ms than on the others. We believe that such loss of signal of interest is not a flaw of the SMCS approach itself, and could be prevented by finer tuning of the parameters. Indeed, on the trial presented on [\(b\)](#) and [\(c\)](#), observe that a second, spontaneous neuronal propagation occurs (peaking around 890 ms), with almost no influence on the periodic artifacts component. This last feature is of particular interest to us, regarding our attempt to capture inter-trial variability. Even though the trial was processed for reproducible, evoked response, our method was able to extract a significant neuronal event without prior knowledge of it.

4.3.2. Spontaneous Activity

This last observation motivates the screening of blank acquisitions for spontaneous activity. After the first bleaching approximation on blank acquisitions, we put apart those presenting obvious patterns of spontaneous activity, for two reasons. First, they might bias the estimation of the periodic artifacts amplitude, what will impact the processing of the stimulus condition acquisitions. Second, we can actually apply the SMCS method to extract spontaneous activity from noise and artifacts, at the single trial level.

On [Figure 7](#), [\(e\)](#) looks like a cleaner version of [\(b\)](#) (compare frames 210 and 240 ms), for the hundred milliseconds following the stimulus onset: the spread of activity arise from the same spatial location, with the same delay, and similar amplitude.

In contrast, spontaneous activity can only be observed at the single trial level. As shown in [Figure 8\(b\)](#) the SMCS method performs as well for spontaneous activity as for evoked activity (note again, however, that some neuronal signal has been captured by the periodic artifacts component). On that trial, the activity does not originates from the same locations as the above evoked activity, neither does it occurs at the same time. For

that reason, after averaging over only nine blank acquisitions, individual neuronal events are almost indistinguishable on (e).

Discussion

As expected by considering § 1, the SMCS model appears rather well adapted to VSDOI data. It seems general enough for processing various data sets, obtained with different experimental set-ups, over different subjects, to investigate different phenomenon. Of course, the results presented here should be completed by study from other data sets and by investigating more complex cortical functional properties.

The first results presented here are already informative concerning the use of VSDOI *in-vivo*. It is clear to us that the high frequency fluctuations are dominated by noise, which should be removed prior to fine analysis of VSDOI data. Averaging acquisitions across trials is usually the best way to do it, but this prevents analysis of trial-to-trial variability, as illustrated in § 4.3.2. Moreover, simple linear smoothing of the data is not an option to us, since many sharp features of the signal would be lost. If one is interested in recovering complex transient signals (such as propagating waves), those arguments are in favor of a complex, non-linear separation method such as the SMCS.

However, the difficulty of the problem at hand makes it necessary to enforce as many priors one can have on the signal; this plays against a method aiming to be general. As far as we can tell from the above results, the wavelet sparsity of the signal is a good prior for VSDOI, allowing in particular adaptive removal of the noise. In addition, it can also separate the neuronal signal from periodic artifacts in many instances but this does not seem very robust. This task is actually what renders the whole SMCS method quite involved.

In particular the very definition of the periodic artifacts component is vague, especially since neuronal activity is known to present oscillatory behaviors at various frequencies. In addition, important information over those artifacts is learnt from the blank acquisitions, which might themselves contains important neuronal signals, which would in turn be considered as artifacts later on. For those reasons, we insist on the need of our method for clean blank acquisitions, and that synchronizing acquisitions with heartbeat and respiration is useful. Note that the real data tested here are recorded on anesthetized animals, and this last inconvenient might reveal worst in acquisitions on awake animals.

Finer study of what actually lies in the periodic artifacts (notably according to the frequency range) should also be considered. This could be done either simply by looking for spatio-temporal patterns (as veins over the field of view) or by testing for functional bias (as we did with orientation selectivity, see Figure 6).

Acknowledgment

This work was supported by Centre National de la Recherche Scientifique (France), the European Research Council (ERC project SIGMA-Vision), the European Union Seventh Framework Programme BrainScaleS (FP7-ICT-2009-6, N 269921),

Brain-i-Nets (FP7-ICT-2009-C, 243914), the Agence Nationale pour la Recherche (SensoryProcessing, Transtact).

We are very grateful to Laurent Demanet for providing us access to the *Wave* computing server at the Massachusetts Institute of Technology.

Bibliography

- L. B. Cohen, B. M. Salzberg, H. V. Davila, W. N. Ross, D. Landowne, A. S. Waggoner, C. H. Wang, Changes in axon fluorescence during activity: molecular probes of membrane potential, *Journal of Membrane Biology* 19 (1) (1974) 1–36.
- H. V. Davila, B. M. Salzberg, L. B. Cohen, A. S. Waggoner, A Large Change in Axon Fluorescence that Provides a Promising Method for Measuring Membrane Potential, *Nature New Biology* 241 (109) (1973) 159–160.
- B. M. Salzberg, H. V. Davila, L. B. Cohen, Optical Recording of Impulses in Individual Neurons of an Invertebrate Central Nervous System, *Nature* 246 (5434) (1973) 508–509.
- A. Grinvald, R. Hildesheim, VSDI: a new era in functional imaging of cortical dynamics, *Nature Reviews Neuroscience* 5 (11) (2004) 874–885.
- A. Grinvald, D. Shoham, A. Shmuel, D. Glaser, I. Vanzetta, E. Shtoyerman, H. Slovlin, C. Wijnbergen, R. Hildesheim, A. Arieli, In-vivo optical imaging of cortical architecture and dynamics, in: *Modern Techniques in Neuroscience Research*, Springer, 893–969, 1999.
- R. D. Frostig (Ed.), *In Vivo Optical Imaging of Brain Function*, Frontiers in Neuroscience, CRC Press, 2nd edn., 2009.
- G. J. Foschini, R. D. Gitlin, J. Salz, Optimum direct detection for digital fiber-optic communication systems, *The Bell System Technical Journal* 54 (8) (1975) 1389–1430.
- D. S. Peterka, H. Takahashi, R. Yuste, Imaging voltage in neurons, *Neuron* 69 (1) (2011) 9–21.
- K. Holthoff, D. Zecevic, A. Konnerth, Rapid time course of action potentials in spines and remote dendrites of mouse visual cortex neurons, *The Journal of Physiology* 588 (7) (2010) 1085–1096.
- C. D. Acker, P. Yan, L. M. Loew, Single-Voxel Recording of Voltage Transients in Dendritic Spines, *Biophysical Journal* 101 (2) (2011) L11–L13.
- I. Ferezou, S. Bolea, C. C. H. Petersen, Visualizing the Cortical Representation of Whisker Touch: Voltage-Sensitive Dye Imaging in Freely Moving Mice, *Neuron* 50 (4) (2006) 617–629.
- M. T. Lippert, K. Takagaki, W. Xu, X. Huang, J. Y. Wu, Methods for voltage-sensitive dye imaging of rat cortical activity with high signal-to-noise ratio, *Journal of Neurophysiology* 98 (1) (2007) 502–512.
- S. Chemla, F. Chavane, Voltage-sensitive dye imaging: Technique review and models, *Journal of Physiology Paris* 104 (1-2) (2010a) 40–50.
- T. Tominaga, Y. Tominaga, A new non-scanning confocal microscopy module for functional voltage-sensitive dye and calcium imaging of neuronal circuit activity, *Journal of Neurophysiology* 110 (2) (2013) 553–561.
- S. Chemla, F. Chavane, A biophysical cortical column model to study the multi-component origin of the VSDI signal, *NeuroImage* 53 (2) (2010b) 420–438.
- A. Arieli, D. Shoham, R. Hildesheim, A. Grinvald, Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex., *Journal of Neurophysiology* 73 (5) (1995a) 2072–2093.
- L. Muller, A. Reynaud, F. Chavane, A. Destexhe, The stimulus-evoked population response in visual cortex of awake monkey is a propagating wave, *Nature Communications* 5 (2014) 3675.
- A. Arieli, A. Sterkin, A. Grinvald, A. Aertsen, Dynamics of Ongoing Activity: Explanation of the Large Variability in Evoked Cortical Responses, *Science* 273 (5283) (1996) 1868–1871.
- I. Lampl, I. Reichova, D. Ferster, Synchronous Membrane Potential Fluctuations in Neurons of the Cat Visual Cortex, *Neuron* 22 (2) (1999) 361–374.
- C. C. H. Petersen, T. T. G. Hahn, M. Mehta, A. Grinvald, B. Sakmann, Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex, *Proceedings of the National Academy of Sciences* 100 (23) (2003) 13638–13643.
- D. Contreras, Propagating waves in visual cortex, *Neuron* 55 (1) (2007) 3–5.
- W.-F. Xu, X.-Y. Huang, K. Takagaki, J.-Y. Wu, Compression and reflection of visually evoked cortical waves, *Neuron* 55 (2007) 119–129.
- A. Benucci, R. A. Frazor, M. Carandini, Standing waves and traveling waves distinguish two circuits in visual cortex, *Neuron* 55 (1) (2007) 103–117.

- L. Muller, A. Destexhe, Propagating waves in thalamus, cortex and the thalamocortical system: Experiments and models, *Journal of Physiology Paris* 106 (5-6) (2012) 222–238.
- S. E. Boustani, O. Marre, S. Béhuret, P. Baudot, P. Yger, T. Bal, A. Destexhe, Y. Frégnac, Network-state modulation of power-law frequency-scaling in visual cortical neurons, *PLoS Comp. Biol.* 5 (9) (2009) e1000519.
- H. J. Tanke, P. van Oostveldt, P. van Duijn, A parameter for the distribution of fluorophores in cells derived from measurements of inner filter effect and reabsorption phenomenon, *Cytometry* 2 (6) (1982) 359–369.
- L. Song, E. J. Hennink, I. T. Young, H. J. Tanke, Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy, *Biophysical Journal* 68 (6) (1995) 2588–2600.
- Y. Chen, W. S. Geisler, E. Seidemann, Optimal temporal decoding of neural population responses in a reaction-time visual detection task, *Journal of Neurophysiology* 99 (3) (2008) 1366–1379.
- A. Reynaud, S. Takerkart, G. S. Masson, F. Chavane, Linear model decomposition for voltage-sensitive dye imaging signals: application in awake behaving monkey, *NeuroImage* 54 (2) (2011) 1196–1210.
- K. Takagaki, M. T. Lippert, B. Dann, T. Wanger, F. W. Ohl, Normalization of Voltage-Sensitive Dye Signal with Functional Activity Measures, *PLoS ONE* 3 (12) (2008) e4041.
- D. Shoham, D. E. Glaser, A. Arieli, T. Kenet, C. Wijnbergen, Y. Toledo, R. Hildesheim, A. Grinvald, Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes, *Neuron* 24 (4) (1999) 791–802.
- T. Bonhoeffer, A. Grinvald, Optical imaging based on intrinsic signals: the methodology, in: A. Toga, J. Mazziotta (Eds.), *Brain Mapping: The Methods*, Academic Press, San Diego, CA, 55–97, 1996.
- V. Markounikau, C. Igel, A. Grinvald, D. Jancke, A Dynamic Neural Field Model of Mesoscopic Cortical Activity Captured with Voltage-Sensitive Dye Imaging, *PLoS Computational Biology* 6 (9) (2010) e1000919.
- D. Jancke, F. Chavane, S. Naaman, A. Grinvald, Imaging cortical correlates of illusion in early visual cortex, *Nature* 428 (6981) (2004) 423–426.
- S. Chakraborty, A. Sandberg, S. A. Greenfield, Differential dynamics of transient neuronal assemblies in visual compared to auditory cortex, *Experimental Brain Research* 182 (4) (2007) 491–498.
- A. Arieli, A. Sterkin, A. Grinvald, A. A., Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses, *Science* 273 (1995b) 1868–1871.
- H. T. Ma, C. H. Wu, J. Y. Wu, Initiation of spontaneous epileptiform events in the rat neocortex in vivo, *Journal of Neurophysiology* 91 (2) (2004) 934–945.
- I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2002.
- A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* 13 (4-5) (2000) 411–430.
- S. Maeda, S. Inagaki, H. Kawaguchi, W. J. Song, Separation of signal and noise from in vivo optical recording in Guinea pigs using independent component analysis, *Neuroscience Letters* 302 (2-3) (2001) 137–140.
- S. Inagaki, T. Katura, H. Kawaguchi, W. J. Song, Isolation of neural activities from respiratory and heartbeat noises for in vivo optical recording in guinea pigs using independent component analysis, *Neuroscience Letters* 352 (1) (2003) 9–12.
- J. Reidl, J. Starke, D. B. Omer, A. Grinvald, H. Spors, Independent component analysis of high-resolution imaging data identifies distinct functional domains, *NeuroImage* 34 (1) (2007) 94–108.
- S. Onat, P. König, D. Jancke, Natural scene evoked population dynamics across cat primary visual cortex captured with voltage-sensitive dye imaging, *Cerebral Cortex* 21 (11) (2011a) 2542–2554.
- S. Onat, N. Nortmann, S. Rekauszke, P. König, D. Jancke, Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging, *NeuroImage* 55 (4) (2011b) 1763–1770.
- K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. Frith, R. S. J. Frackowiak, Statistical Parametric Maps in Functional Imaging: A General Linear Approach, *Human Brain Mapping* 2 (4) (1995) 189–210.
- D. B. Omer, R. Hildesheim, A. Grinvald, Temporally-structured acquisition of multidimensional optical imaging data facilitates visualization of elusive cortical representations in the behaving monkey, *NeuroImage* 82 (2013) 237–251.
- E. Yavuz, Source separation analysis of visual cortical dynamics revealed by voltage sensitive dye imaging, Ph.D. thesis, École doctorale Cerveau, Cognition et Comportement (Paris), 2012.
- S. Mallat, S. Zhong, Characterization of signals from multiscale edges, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 14 (7) (1992) 710–732.
- E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing Sparsity by Reweighted ℓ_1 Minimization, *Journal of Fourier Analysis and Applications* 14 (5) (2008) 877–905.
- H. Raguét, A Signal Processing Approach to Voltage-Sensitive Dye Optical Imaging, Ph.D. thesis, Université Paris-Dauphine, 2014.
- J. Y. Wu, X. Huang, C. Zhang, Propagating Waves of Activity in the Neocortex: What They Are, What They Do, *The Neuroscientist* 14 (5) (2008) 487–502.
- P. L. Combettes, J.-C. Pesquet, Proximal Splitting Methods in Signal Processing, vol. 49 of *Optimization and Its Applications*, chap. 10, Springer, New York, NY, USA, 185–212, 2011.
- H. Raguét, J. M. Fadili, G. Peyré, A Generalized Forward-Backward Splitting, *SIAM Journal on Imaging Sciences* 6 (3) (2013) 1199–1226.
- S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 3rd edn., 2008.
- D. H. Hubel, T. N. Wiesel, Receptive fields of single neurons in the cat's striate cortex, *The Journal of Physiology* 148 (3) (1959) 574–591.
- V. B. Mountcastle, The columnar organization of the neocortex, *Brain* 120 (4) (1997) 701–722.
- D. H. Hubel, T. N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J Physiol.* 160 (1) (1962) 106–154.
- T. Bonhoeffer, A. Grinvald, The Layout of Iso-orientation Domains in Area 18 of Cat Visual Cortex: Optical Imaging Reveals a Pinwheel-like Organization, *The Journal of Neuroscience* 13 (10) (1993) 4157–4180.
- D. Sharon, A. Grinvald, Dynamics and Constancy in Cortical Spatiotemporal Patterns of Orientation Processing, *Science* 295 (5554) (2002) 512–515.
- F. Chavane, D. Sharon, D. Jancke, O. Marre, Y. Frégnac, A. Grinvald, Lateral spread of orientation selectivity in V1 is controlled by intracortical cooperativity, *Frontiers in Systems Neuroscience* 5 (4) (2011) 1–26.
- G. G. Blasdel, G. Salama, Voltage-sensitive dyes reveal a modular organization in monkey striate cortex, *Nature* 321 (6070) (1986) 579–585.

Figure 7, on page 25, and Figure 8, on page 26: Component separation results for Mouse_A11, for stimulus condition and spontaneous activity, respectively. The time courses correspond to spatial positions indicated on the frames by a square of the same color. Vertical dashed line indicates stimulus onset, when present. (a) Observations normalized by the estimated gain, single trial. (b) Neuronal response, single trial. (c) Periodic artifacts, single trial. (d) Residual, single trials. (e) Neuronal response, average across trials, single trials in transparency.

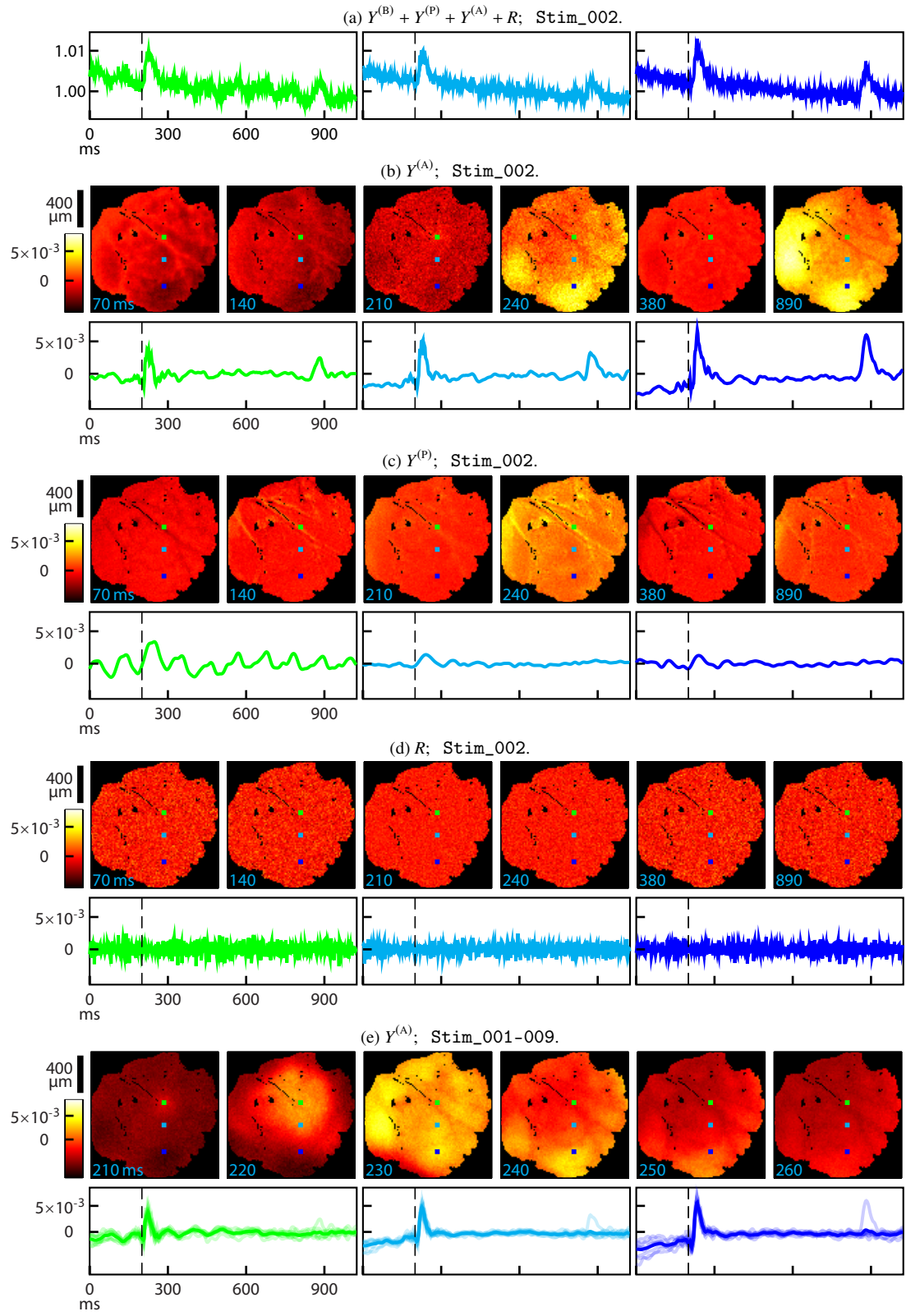


Figure 7: Mouse_A11 data, stimulus conditions. See *opposite*.

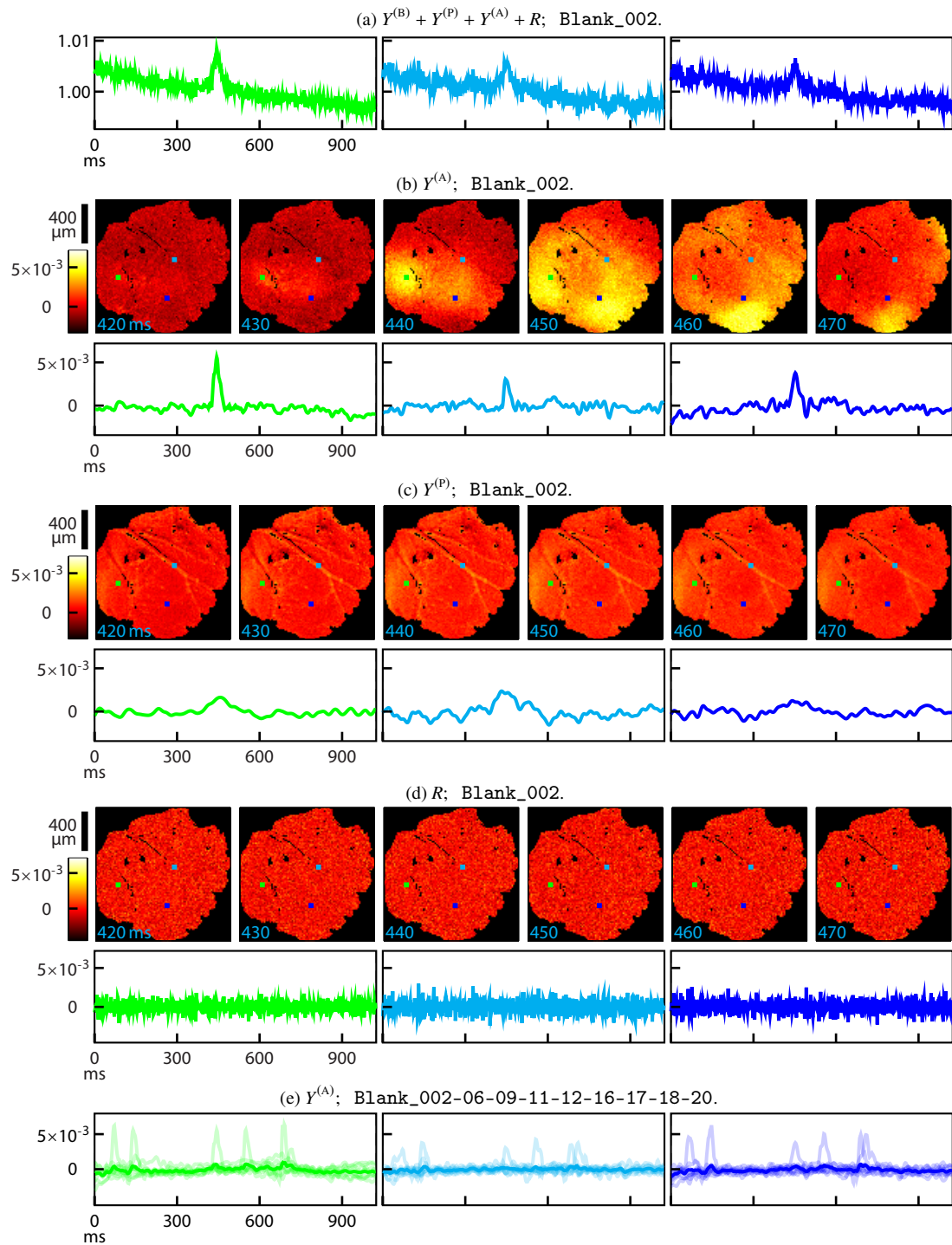


Figure 8: Mouse_All data, spontaneous activity. See bottom of [page 24](#).